TRANSACTIONS

OF THE

AMERICAN INSTITUTE

OF

ELECTRICAL ENGINEERS

VOLUME 77

1958

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1959 by the American Institute of Electrical Engineers

Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1958 (volume 77) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

- Aircraft and Flight Test Instrumentation
- Basic Sciences
- Communication Switching Systems
- Communication Theory
- Computing Devices
- Data Communication
- Electrical Insulation
- Electrical Techniques in Medicine and Biology
- Electronic and High Frequency Instruments
- Electronics
- Indicating and Integrating Instruments
- Magnetic Amplifiers
- Nucleonic and Radiation Instrumentation
- Nucleonics
- Radio Communication Systems
- Recording and Controlling Instruments
- Semiconductor Metallic Rectifiers
- Solid State Devices
- Special Instruments and Auxiliary Apparatus
- Telegraph Systems
- Telemetering
- Television and Aural Broadcasting Systems
- Wire Communication Systems

Part II. Applications and Industry

- Air Transportation
- Chemical Industry
- Domestic and Commercial Applications
- Electric Heating
- Electric Welding
- Feedback Control Systems
- General Industry Applications
- Industrial and Commercial Power Systems
- Industrial Control
- Industrial Power Rectifiers
- Land Transportation
- Marine Transportation
- Metal Industry
- Mining Industry
- Petroleum Industry
- Production and Application of Light

Part III. Power Apparatus and Systems

- Insulated Conductors
- Power Generation
- Power System Communications
- Protective Devices
- Relays
- Rotating Machinery
- Substations
- Switchgear
- System Engineering
- Transformers
- Transmission and Distribution

- Education
- Research
- Safety

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 77 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., February 2–7, 1958.
2. Middle Eastern District Meeting, Washington, D. C., April 28–30, 1958.
3. Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958.
4. East Central District Meeting, Huntington, W. Va., May 13–15, 1958.
5. Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–26, 1958.
6. Pacific General Meeting, Sacramento, Calif., August 19–22, 1958.
7. Fall General Meeting, Pittsburgh, Pa., October 13–15, 1958.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Vol. 7
Pt. 1
1958

Communication and Electronics Index for 1958

1. Technical Subject Index

Abnormal Corona, Some Measurements of. Penney, Hewitt....319–27
A-C Computers, A Differentiator for. Johnson....1–4
A-C Controlled Magnetic Amplifiers. Lehtonen, Cronauer....476–80
Accuracy and Speed in Analog-to-Digital Transducer Measurements, Relay Scanning Design Technique Generates High. Kay....248–50
Accurate Determination of Capacitance by Conjoining Analytical and Analog Techniques. Horgan, Pesavento....397–400
Accurate to 0.1 Per Cent, An Analog Frequency Measuring Circuit. Mitchell....983–5; disc. 985
(Air Force) Automatic Telegraph Switching System Plan 55-A. Vernam....239–47
Air–Ground Telephone Service Trial, Public. Augustus....926–30
Airborne Teletypewriter AN/AGC-1. Michals..234–9
Amplifier, A Stabilized D-C Differential Transistor. Depian, Smith....157–9
Amplifier as a Laboratory Tool, The Operational. Pfeiffer....947–51; disc. 951
Amplifiers Using Dynistors or Trinistors, Linear Power. Hierholzer....892–8
Amplifiers with Random Noise Signals, Performance of Class B Audio. Usher....939–43
Amplitude Random Function Generator, A Constant-. Hellwarth....443–52
Amplitude-Stabilized Bridged-T Oscillator, An. Enslein....75–9
AN/AGC-1, Airborne Teletypewriter. Michals..234–9
Analog Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems, An. Elgerd....358–66; disc. 366
Analog Computer Study of the Transient Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant. Leiby....17–25
Analog Computers, Matrix Programming of Electronic. Horn, Honnell....420–6; disc. 426
Analog Frequency Measuring Circuit Accurate to 0.1 Per Cent, An. Mitchell....983–5; disc. 985
Analog Techniques, Accurate Determination of Capacitance by Conjoining Analytical and. Horgan, Pesavento....397–400
Analog-to-Digital Transducer Measurements, Relay Scanning Design Technique Generates High Accuracy and Speed in. Kay....248–50
Analysis Applied to the Saturable Reactor, Block-Diagram Method of. Hubbard....57–64; disc. 65
Analysis of Logical Networks, Matrix. Schubert....10–13
Analysis of Nonresonant Charging Circuit for Magnetic Pulse, Generator. Weinberg....271–7
Analysis of Redundancy Networks, The. Moskowitz....627–32
Analytical and Analog Techniques, Accurate Determination of Capacitance by Conjoining. Horgan, Pesavento....397–400
Antenna, A Quadruplexer Allowing the Simultaneous Transmission of Two Complete Television Stations Using a Common. MacKimmie....787–91
Antenna System, A New Broad-Band Microwave. Friis, May....97–100
Apparatus, Philosophy of Applying Digital Computers to the Design of Electric. Abetti, Cuthbertson, Williams....367–77; disc. 377
Application of Negative-Impedance Repeaters on Long Rural Telephone Lines, The. Uthlaut....230–4
Application of Zener Diodes to Expanded Scale Instruments, On the. Corson....535–9
Applications of Nonlinear Magnetics. Storm....380–8; disc. 388
Assessment of Effects of Delay Distortion in Data Systems. Fowler, Gibby....918–23
Attenuation at 11 Kmc and Some Implications Affecting Relay System Engineering, Radio. Hathaway, Evans....930–8; disc. 1061
Audio Amplifiers with Random Noise Signals, Performance of Class B. Usher....939–43
Automatic Control of Electrical Precipitation Rectifiers. Van Hoesen, White, Hall....126–8
Automatic Mobile Telephone System. Dimmer..723–6
Automatic Telegraph Switching System Plan 55-A. Vernam....239–47
Avalanche Transistors, Theory of Transient Build-up in. Shockley, Gibbons....993–8

B

Balancing Reactors for Semiconductor Rectifiers, Current-. Dortort....452–6; disc. 456
(Basic Sciences) A Generalization of the Calculus of Finite Differences to Nonuniformly Spaced Variables. Kron....539–44
(Basic Sciences) Accurate Determination of Capacitance by Conjoining Analytical and Analog Techniques. Horgan, Pesavento....397–400
(Basic Sciences) An Electric Circuit Theory Approach to Finite Difference Stability. Karplus....210–13
(Basic Sciences) Certain Applications of Matrices to Circuit Theory. Pipes....251–6
(Basic Sciences) Matrix Analysis of Logical Networks. Schubert....10–13
(Basic Sciences) Matrix Programming of Electronic Analog Computers. Horn, Honnell....420–6; disc. 426
(Basic Sciences) Measurement of the Scattering Matrix with an Intervening Ionosphere. Bryak....611–12
(Basic Sciences) Reliable Stochastic Sequential Switching Circuits. Mullin....606–11
(Basic Sciences) Subharmonics in Iron-Cored Reactor Oscillatory Circuits. Mostafa, El-Karaksi..104–13
(Basic Sciences) The Numerical Evaluation of Expressions Involving Complete Elliptic Integrals. Grover....496–502
Basis for Transmission Performance Objectives in a Telephone Communication System, A. MacAdam....205–08; disc. 208
Behavior of the Ferroresonant Series Circuit Containing a Square-Loop Reactor. Dennard....903–11
B-H Loops on Magnetic Amplifiers, Reactions of Bistable. Morgan....586–90
Bibliography, Magnetic Amplifier—1951-1956. (Committee Report)....613–27
Bibliography, 1957 Magnetic Amplifier. (Committee Report)....1051–7
Binary Code, An Experimental Study of a. Peterson....388–92; disc. 392
Binary Communication Feedback Systems. Harris, Hauptschein, Morgan, Schwartz....960–9
Binary Symmetric Channel, On Coding for the. Fontaine, Peterson....638–46; disc. 646
Binary Symmetric Decision Feedback Systems. Harris, Morgan....436–43
Bistable B-H Loops on Magnetic Amplifiers, Reactions of. Morgan....586–90
Bits of Information. Zamanakos....197–201
Block-Diagram Method of Analysis Applied to the Saturable Reactor. Hubbard....57–64; disc. 65
Block Length for Data Transmission with Error Checking, Optimum. Wood....855–61
Boiling Water Reactor, An Electronic Feedwater Control System for the Experimental. Morse, Brey....679–83
Boiling Water Reactor Nuclear Power Plant, Analog Computer Study of the Transient Performance of a Dual Cycle. Leiby....17–25
Branch Exchanges, A New Small Crossbar Telephone System for Private. Abbott....911–18
Bridged-T Oscillator, An Amplitude-Stabilized. Enslein....75–9
Broad-Band Microwave Antenna System, A New. Friis, May....97–100
Broadcasting, Video Testing Techniques in Television. Ste-Marie....26–33
Build-up in Avalanche Transistors, Theory of Transient. Shockley, Gibbons....993–8

C

Cable Carrier System, A Transistorized Repeater for Use with the 45BN. Babin, Fish....41–9
Cable Carrier System, Determination of the Least First Cost of a. Krzyczkowski....778–82; disc. 782
Cable Telegraphy, An Improved 6-Channel Time Division Multiplex for Submarine. Wilder....217–22
Cables and Methods Taken to Improve Them, Study of Return Losses on Loaded Trunk. Speake....201–04; disc. 204
Cables, Surge Measurement Errors Introduced by Coaxial. Park....343–7; disc. 347
Calculation of Network Functions Used in Loss Formula Studies, Digital. Ferguson, Long, Rindt....647–52; disc. 652
Calculation of Transient Performance of the Primary Coolant System in a Water Reactor, Digital. Lewis....1036–43
Calculus of Finite Differences to Nonuniformly Spaced Variables, A Generalization of the. Kron..539–44
Canada TD-2 System, The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-. Curtis, Strahlendorf, Wade....185–90
Capacitance by Conjoining Analytical and Analog Techniques, Accurate Determination of. Horgan, Pesavento....397–400
Carrier Repeater for Military Use, A Frequency-Converting Telephone. Goltsos, Johnston, Anderson....432–6
Carrier System, A Transistorized Repeater for Use with the 45BN Cable. Babin, Fish....41–9
Carrier System, Determination of the Least First Cost of a Cable. Krzyczkowski....778–82; disc. 782
Carrier System, Insertion Loss Filter Design Applied to Transistorized. Winkler....519–24
Carrier System, The ON/K. Masters....806–10
Carrier Systems, A New Method of Frequency Error Correction in. Stachiewicz....175–80
Carrier Telephone System, Type OJ. Coy, Kelsey, Nicholls....810–17
Central Station Power, Organic-Moderated Reactors for. Parkins, Weisner....985–93
Certain Applications of Matrices to Circuit Theory. Pipes....251–6
Channel, On Coding for the Binary Symmetric. Fontaine, Peterson....638–46; disc. 646
Channel on the Trans-Canada TD-2 System, The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave. Curtis, Strahlendorf, Wade....185–90
Channels for SAGE Data Systems, Communication. James....838–43
Characteristics of Magnetic Amplifiers Utilizing Square-Loop Core Material and Square-Wave Supply Voltage. Britten....590–7; disc. 597
Characteristics of Serial-Type Digital Data Systems, An Analog Computer Study of the Transient Behavior and Stability. Elgerd....358–66; disc. 366
Charging Circuit for Magnetic Pulse Generator, Analysis of Nonresonant. Weinberg....271–7
Chesapeake Bay, Dial Telephone Service for Smith Island, an Isolated Community in. Littleton....923–6
Circuit, A Magnetic Amplifier Commutating and Pulse-Width Encoding. Lucke....884–92
Circuit Accurate to 0.1 Per Cent, An Analog Frequency-Measuring. Mitchell....983–5; disc. 985
Circuit Containing a Square-Loop Reactor, Behavior of the Ferroresonant. Dennard....903–11
Circuit for Magnetic Pulse Generator, Analysis of Nonresonant Charging. Weinberg....271–7
Circuit, The Ferroresonant. Kelly....843–8; disc. 1061
Circuit Theory Approach to Finite Difference Stability, An Electric. Karplus....210–13
Circuit Theory, Certain Applications of Matrices to. Pipes....251–6
Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint, Dial Switching of Toll. Petrie....481–5
Circuits, Minimization of Components in Electronic Switching. Beatson....283–91; disc. 291
Circuits, Reliable Stochastic Sequential Switching. Mullin....606–11
Circuits, Subharmonics in Iron-Cored Reactor Oscillatory. Mostafa, El-Karaksi....104–13
Class B Audio Amplifiers with Random Noise Signals, Performance of. Usher....939–43
Clocks for Data Transmission, Synchronized. Edson, Flavin, Perry....832–6
Coaxial Cables, Surge Measurement Errors Introduced by. Park....343–7; disc. 347
Code, An Experimental Study of a Binary. Peterson....388–92; disc. 392
Code Construction, Diagrammatic Methods of. Clark....817–23
(Coding) Bits of Information. Zamanakos....197–201
Coding for the Binary Symmetric Channel, On. Fontaine, Peterson....638–46; disc. 646
(Committee Report) Magnetic Amplifier Bibliography—1951–1956. (58-996)....613–27
(Committee Report) 1957 Magnetic Amplifier Bibliography. (58-1328)....1051–7
(Committee Report) Proposed Standard Terms and Definitions for Magnetic Amplifiers. (58-175)....429–31; disc. 431
(Committee Report) Proposed Standards for Core Test Methods for Toroidal Magnetic Amplifier Cores. (58-71)....524–31

(Committee Report) The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage. (58-27)......350–7; disc. 357
Communication Channels for SAGE Data Systems. James..............................838–43
Communication Feedback Systems, Binary. Harris, Hauptschein, Morgan, Schwartz............960–9
Communication Services, Radar Interference to Microwave. Campbell..............717–22; disc. 722
Communication, Statistical Encoding for Text and Picture. Michel..............................33–6
(Communication Switching Systems) A Magnetic-Drum Storage System Considered for Use as a Common Sender In Nation-Wide Dialing. May......5–10
(Communication Switching Systems) A New Small Crossbar Telephone System for Private Branch Exchanges. Abbott.......................911–18
(Communication Switching Systems) Automatic Mobile Telephone System. Dimmer..............723–6
(Communication Switching Systems) Mechanization of Toll Switching at Providence, R.I. Anderson, Thornton..............................954–60
(Communication Switching Systems) Nationwide Switching of Intertoll Trunks—Transmission Considerations. Wallman.......223–9; disc. 229
(Communication Switching Systems) Optimal Utilization and Extension of Interoffice Trunking Facilities. Kalaba, Juncosa.............998–1002; disc. 1002
Communication System, A Basis for Transmission Performance Objectives in a Telephone. MacAdam.................................205–08; disc. 208
(Communication Theory) An Experimental Study of a Binary Code. Peterson.........388–92; disc. 392
(Communication Theory) Binary Symmetric Decision Feedback Systems. Harris, Morgan.......436–43
(Communication Theory) Bits of Information. Zamanakos....................................197–201
(Communication Theory) On Coding for the Binary Symmetric Channel. Fontaine, Peterson.................................638–46; disc. 646
(Communication Theory) Recent Progress in Applying Information Theory to Digital Transmission Systems. Filipowsky........................848–55
(Communication Theory) The Analysis of Redundancy Networks. Moskowitz.....................627–32
Communications, Operations Research in. Harris, Hauptschein, Schwartz...........633–7; disc. 637
Communications System for the Massachusetts Turnpike, An Integrated Mobile Microwave. Lowe..............................703–06; disc. 706
Commutating and Pulse-Width Encoding Circuit, A Magnetic-Amplifier. Lucke..............884–92
Compandor, A Transistorized. Perkins, Ferreault, Perkins..................................791–7
Component Parts Slection, The Economics of. Fuegel.....................................13–16
Components for High Storage Density, Magnetic Drum. Hong..................................667–72
Components in Electronic Switching Circuits, Minimization of. Beatson..............283–91; disc. 291
Comprehensive Study of Magnetic-Amplifier Design, A. Hubbard, Bishop:
  Part I..................................562–72
  Part II.......................572–80; disc. 580
Computer, A General Approach for Obtaining Transient Response by the Use of a Digital. Lego, Sze..1031–6
Computer Aids in Power Pool Operation Studies, Digital. McIntyre, Blake, Clubb.652–7; disc. 657
Computer in a Generator Reserve Requirement Study, The Use of a Digital. Brown...............82–5
Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems, An Analog. Elgerd....358–66; disc. 366
Computer Study of the Transient Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant, Analog. Leiby..........................17–25
Computer Vertification of Dynamic Action, Steam Generator Instrumentation for a Nuclear Power Plant with Analog. Lynch, Waite.........690–6
Computers, A Differentiator for A-C. Johnson....1–4
Computers, Direct System Synthesis by Means of. Clymer...............................798–806
Computers, Function Tables in Digital Control. Shubert...................................316–19
Computers, Matrix Programming of Electronic Analog. Horn, Honnell..................420–6; disc. 426
Computers to Engineering Problems, The Practice and the Economics of Applying Digital. Abetti, Wiliams.......................331–41; disc. 341
Computers to the Design of Electric Apparatus, Philosophy of Applying Digital. Abetti, Cuthbertson, Williams.....................367–77; disc. 377
(Computing Devices) Digital Calculation of Network Functions Used in Loss Formula Studies. Ferguson, Long, Rindt.....................647–52; disc. 652
(Computing Devices) Digital Calculation of Transient Performance of the Primary Coolant System in a Water Reactor. Lewis.................1036–43
(Computing Devices) Logic for Applying Topological Methods to Electric Networks. Byerly, Long, King.............................657–65; disc. 665
(Computing Devices) Magnetic Drum Components for High Storage Density. Hong............667–72
(Computing Devices) Relay Scanning Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements. Kay..248–50
(Computing Devices) Shifting Counters. Eldert, Gray, Gurk, Rubinoff............................70–4
(Computing Devices) Some Practical Mathematical Techniques. Carter..................295–302
(Computing Devices) Spectrum Analysis of Random Noise Generators. Polimerou.............672–9
(Computing Devices) The Design of Function Generators Using Short-Time Memory Devices and Nonlinear Elements. Revay, Ford............143–52
(Computing Devices) The Economics of Component Parts Selection. Fuegel..................13–16
(Computing Devices) The Operational Amplifier as a Laboratory Tool. Pfeiffer......947–51; disc. 951
Concepts in Reactor Control Instrumentation, Modular. Cockrell, Ryan, Magee..................696–9
Conductors, Magnetic Field Strength in Oersteds Near Large Rectangular. Miller......726–7; disc. 727
Considerations, SAGE Data System. Enticknap, Schuster..................................824–32
Constant-Amplitude Random Function Generator, A. Hellwarth..............................443–52
Construction, Diagrammatic Methods of Code. Clark...817–23
Control Amplifier Using a Saturable Current Transformer and a Switching Transistor, A New. Morgan.........................557–62; disc. 562
Control Circuit Inductance, Elimination of Magnetic Amplifier. Chandler, Downing..........................550–6; disc. 556
Control of Electrical Precipitation Rectifiers, Automatic. VanHoesen, White, Hall...........126–8
Controlled Magnetic Amplifiers, A-C. Lehtonen, Cronauer..................................476–80
Controlled Rectifier: Key to the Continuing Control Renaissance, The. Harnden..1006–12; disc. 1012
Controller Using a Magnetic Modulator, A Triple-Action Positioning. Bullen...............785–7
Converters, The Definition and Measurement of the Time Constant and Response Time of Thermal. Hermach ....277–82; disc. 282
Converting Telephone Carrier Repeater for Military Use, A Frequency-. Goltsos, Johnston, Anderson 423–6
Coolant Instrumentation for High-Performance Pressurized Water Reactor Power Plants, Main. Noyes 683–90
Coolant System in a Water Reactor, Digital Calculation of Transient Performance of the Primary. Lewis...1036–43
Cooling Fins, Thermal Impedance of. Diebold, Luft.............................739–45; disc. 745
Core Material and Square-Wave Supply Voltage, Characteristics of Magnetic Amplifiers Utilizing Square-Loop. Britten.................590–7; disc. 597
Core-Reset Functions in Magnetic-Amplifier Analysis. Feth:
  Part I. A Core-Reset Function....503–09; disc. 509
  Part II. Determination of Amplifier Performance..510–19
Core Test Methods for Toroidal Magnetic Amplifier Cores, Proposed Standards for. (Committee Report).....................................524–31
Core Timing Device, Magnetic Saturable-. Lowrance 393–7
Corona, Some Measurements of Abnormal. Penney, Hewitt...................................319–27
Correction in Carrier Systems, A New Method of Frequency Error. Stachiewicz..............175–80
Cost of a Cable Carrier System, Determination of the Least First. Krzyczkowski......778–82; disc. 782
Counters, Shifting. Eldert, Gray, Gurk, Rubinoff.70–4
Criteria for Low-Level Second Harmonic Magnetic Modulators, Design. Kletsky...........1013–19
Crossbar Telephone System for Private Branch Exchanges, A New Small. Abbott...........911–18
Current-Balancing Reactors for Semiconductor Rectifiers. Dortort..................452–6; disc. 456
Current-Limiting Fuse as Fault Protection for Semiconductor Rectifiers, The. Gutzwiller...............................751–4; disc. 754
Current Transformer and a Switching Transistor, A New Control Amplifier Using a Saturable. Morgan..............................557–62; disc. 562
Cycle Boiling Water Reactor Nuclear Power Plant, Analog Computer Study of the Transient Performance of a Dual. Leiby.......................17–25

D

(Data Communication) Error-Checking for 5-Channel Telegraphic Tape. Barbeau............190–3
Data System Considerations, SAGE. Enticknap, Schuster..................................824–32
Data Systems, An Analog Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type Digital. Elgerd..........358–66; disc. 366
Data Systems, Assessment of Effects of Delay Distortion in. Fowler, Gibby.......................918–23
Data Systems, Communication Channels for SAGE. James..................................838–43
Data Terminals, SAGE. Soffel, Spack...........872–9
Data Transmission Over Telephone Lines, A Frequency Modulation Digital Subset for. Weber....867–72
Data Transmission, Synchronized Clocks for. Edson, Flavin, Perry...........................832–6
Data Transmission Testing Set. Boughtwood, Christie...101–04
Data Transmission with Error Checking, Optimum Block Length for. Wood................855–61
D-C Differential Transistor Amplifier, A Stabilized. Depian, Smith...........................157–9
D-C Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors, Reversible-Polarity. Schmitz, Bernstein...........1058–60; disc. 1060
D-C Shunt Motor Drive, High Efficiency Push-Pull Magnetic Amplifier and Its Use as a. Boyajian..545–50
Dead Space, Linearization of. Jones...........836–8
Decision Feedback Systems, Binary Symmetric. Harris, Morgan..................................436–43
Definition and Measurement of the Time Constant and Response Time of Thermal Converters, The. Hermach.......................277–82; disc. 282
Definitions for Magnetic Amplifiers, Proposed Standard Terms and. (Committee Report)..............................429–31; disc. 431
Delay Distortion in Data Systems, Assessment of Effects of. Fowler, Gibby......................918–23
Demand Meter, New Printing. Terry...........80–2
Design, A Comprehensive Study of Magnetic-Amplifier. Hubbard, Bishop:
  Part I..................................562–72
  Part II.......................572–80; disc. 580
Design Criteria for Low-Level Second-Harmonic Magnetic Modulators. Kletsky..............1013–19
Design, Magnetic Amplifier. Anderson:
  Part I. The Core..........................160–7
  Part II. The Windings....................167–75
Design of a Pile Oscillator for the Ford Nuclear Reactor, The. Stone, Kerr......................1003–06
Design of Electrical Apparatus, Philosophy of Applying Digital Computers to the. Abetti, Cuthbertson, Williams.......................367–77; disc. 377
Design of Function Generators Using Short-Time Memory Devices and Nonlinear Elements, The. Revay, Ford..143–52
Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements, Relay Scanning. Kay...................248–50
Determination of Capacitance by Conjoining Analytical and Analog Techniques, Accurate. Horgan, Pesavento................................397–400
Determination of the Least First Cost of a Cable-Carrier System. Krzyczkowski.........778–82; disc. 782
Device, Magnetic Saturable-Core Timing. Lowrence...393–7
Diagrammatic Methods of Code Construction. Clark..817–23
Dial Switching of Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint. Petrie..481–5
Dial Telephone Service for Smith Island, An Isolated Community in Chesapeake Bay. Littleton..923–6
Dialing, A Magnetic-Drum Storage System Considered for Use as a Common Sender in Nation-Wide. May...5–10
Difference Stability, An Electric Circuit Theory Approach to Finite. Karplus...............210–13
Differences to Nonuniformly Spaced Variables, A Generalization of the Calculus of Finite. Kron..539–44
Differential Transistor Amplifier, A Stabilized D-C. Depian, Smith...........................157–9
Differentiator for A-C Computers, A. Johnson....1–4
Diffused Structures, Silicon Controlled Rectifiers from Oxide Masked. Aldrich, Holonyak........952–4
Digital and Pictorial Photographic Electronic Recorder. McPherson, Sonderby..........................194–6
Digital Calculation of Network Functions Used in Loss Formula Studies. Ferguson, Long, Rindt.............................647–52; disc. 652
Digital Calculation of Transient Performance of the Primary Coolant System in a Water Reactor. Lewis...1036–43
Digital Computer, A General Approach for Obtaining Transient Response by the Use of a. Lego, Sze...1031–6
Digital Computer Aids in Power Pool Operation Studies. McIntyre, Blake, Clubb.........652–7; disc. 657
Digital Computer in a Generator Reserve Requirement Study, The Use of a. Brown................82–5
Digital Computers to Engineering Problems, The Practice and the Economics of Applying. Abetti, Williams.....................331–41; disc. 341
Digital Computers to the Design of Electric Apparatus, Philosophy of Applying. Abetti, Cuthbertson, Williams.....................367–77; disc. 377
Digital Control Computers, Function Tables in. Schubert.....................................316–19

Digital Data Systems, An Analog Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type. Elgerd..........358–66; disc. 366
Digital Indicator with Elpak Translation Logic, An Electroluminescent. Sack..............113–18
Digital Subset for Data Transmission Over Telephone Lines, A Frequency Modulation. Weber..867–72
Digital Transmission Systems, Recent Progress in Applying Information Theory to. Filipowsky....848–55
Diode Reactance Modulator, The. Montgomery.980–3
Diodes to Expanded Scale Instruments, On the Application of Zener. Corson....................535–9
Direct Coupled Transistor Servo Preamplifier, A Stable. DeSautels..........................943–7
Direct System Synthesis by Means of Computers. Clymer..........................798–806
Disk Random Selectors, Spinning. Roberts....755–60
Distortion in Data Systems, Assessment of Effects of Delay. Fowler, Gibby..................918–23
Dividers, Transient Response of Impulse Voltage. Fisher......................411–19; disc. 419
Drive, Transistor-Oscillator Induction-Motor. Card.................................531–5
Drum Components for High Storage Density, Magnetic. Hong..........................667–72
Drum Storage System Considered for Use as a Common Sender in Nation-Wide Dialing, A Magnetic-. May..................................5–10
Dual Cycle Boiling Water Reactor Nuclear Power Plant, Analog Computer Study of the Transient Performance of a. Leiby........................17–25
Dynistors or Trinistors, Linear Power Amplifiers Using. Hierholzer........................ 892–8

E

Economics of Applying Digital Computers to Engineering Problems, The Practice and the. Abetti, Williams......................331–41; disc. 341
Economics of Component Parts Selection, The. Fuegel 13–16
(Education) Television Systems for In-School Teaching. Kraus..............................406–11
Effective Feedback Factor of Self-Balancing Magnetic Amplifiers, The. Geyger..................66–9
Effects of Delay Distortion in Data Systems, Assessment of. Fowler, Gibby..................918–23
Efficiency in Fast-Response Magnetic Amplifiers, Volt-Second Transfer. Pula:
  Part I..................................861–7
Electric Apparatus, Philosophy of Applying Digital Computers to the Design of. Abetti, Cuthbertson, Williams....................367–77; disc. 377
Electric Circuit Theory Approach to Finite Difference Stability, An. Karplus..................210–13
Electric Hygrometer. Meiklejohn............302–05
Electric Networks, Logic for Applying Topological Methods to. Byerly, Long, King.............. ....657–65; disc. 665
Electrical Features of Indian Point Nuclear Electric Generating Station. Reimers............879–84
Electrical Precipitation Rectifiers, Automatic Control of. Van Hoesen, White, Hall................126–8
Electroluminescent Digital Indicator With Elpak Translation Logic, An. Sack..................113–18
Electronic Analog Computers, Matrix Programming of. Horn, Honnell..................420–6; disc. 426
Electronic Feedwater Control System for the Experimental Boiling Water Reactor, An. Morse, Brey.. 679–83
Electronic Recorder, Digital and Pictorial Photographic. McPherson, Sonderby...........194–6
Electronic Switching Circuits, Minimization of Components in. Beatson...........283–91; disc. 291
Electronic Transformers, An Experimental Study of Magnetic Materials for Use in Ultrahigh-Temperature. Harms..................181–4; disc. 184
(Electronics) A Constant-Amplitude Random Function Generator. Hellwarth..................443–52
(Electronics) A Stable Direct-Coupled Transistor Servo Preamplifier. DeSautels..................943–7
(Electronics) A Stabilized D-C Differential Transistor Amplifier. Depian, Smith................157–9
(Electronics) An Amplitude-Stabilized Bridged-T Oscillator. Enslein..........................75–9
(Electronics) An Analog Frequency Measuring Circuit Accurate to 0.1 Per Cent. Mitchell............ ....983–5; disc. 985
(Electronics) Automatic Control of Electrical Precipitation Rectifiers. Van Hoesen, White, Hall...126–8
(Electronics) Design Criteria for Low-Level Second-Harmonic Magnetic Modulators. Kletsky.1013–19
(Electronics) Performance of Class B Audio Amplifiers with Random Noise Signals. Usher.......939–43
(Electronics) Some Measurements of Abnormal Corona. Penney, Hewitt........................319–27
(Electronics) Spinning Disk Random Selectors. Roberts 755–60
(Electronics) The Diode Reactance Modulator. Montgomery..................................980–3
(Electronics) The Ferroresonant Circuit. Kelly....... ....843–8; disc. 1061
(Electronics) Transistor Phase-Locked Oscillators. Edwards, Golubjatnikov, Brady..........1043–51
Elements, The Design of Function Generators Using Short-Time Memory Devices and Nonlinear. Revay, Ford..........................143–52
11 Kmc and Some Implications Affecting Relay System Engineering, Radio Attenuation at. Hathway, Evans.......................930–8; disc. 1061
Elimination of Magnetic-Amplifier Control Circuit Inductance. Chandler, Downing.......... ....550–6; disc. 556
Elliptic Integrals, The Numerical Evaluation of Expressions Involving Complete. Grover.......496–502
Elpak Translation Logic, An Electroluminescent Digital Indicator with. Sack..................113–18
Encoding Circuit, A Magnetic Amplified Commutating and Pulse-Width. Lucke................884–92
Encoding for Text and Picture Communication, Statistical. Michel..........................33–6
Engineering Problems, The Practice and the Economics of Applying Digital Computers to. Abetti, Williams ....331–41; disc. 341
Engineers to the Appreciation of Magnetic Amplifier Problems, Introducing Young. Finzi......119–26
Engineers, Willamette Valley Radio System for the United States Corps of. Pedersen, Marihart...... 401–06
Equipment, Stand-by Power for Pole-Mounted. Jamieson..................................774–8
Equipment, Wide-Band Ultrahigh-Frequency Over-the-Horizon. Felsenheld, Havstad, Jatlow, LeVine, Pollack..................................86–93
Error-Checking for 5-Channel Telegraphic Tape. Barbeau..................................190–3
Error Checking, Optimum Block Length for Data Transmission with. Wood..............855–61
Error Correction in Carrier Systems, A New Method of Frequency. Stachiewicz................175–80
(Error Detecting) Bits of Information. Zamanakos.... 197–201
Errors Introduced by Coaxial Cables, Surge Measurement. Park..................343–7; disc. 347
Evaluation of Expressions Involving Complete Elliptic Integrals, The Numerical. Grover.......496–502
Evaluation of Single Phase Metering Practices Based on Meter Performance and Load Characteristics. Vanderleck..................727–38; disc. 738
Exchanges, A New Small Crossbar Telephone System for Private Branch. Abbott..............911–18
Expansion of the Pacific Coast Microwave Network, The. Kuck..................................898–903
Experimental Boiling Water Reactor, An Electronic Feedwater Control System for the. Morse, Brey.. 679–83
Experimental Study of a Binary Code, An. Peterson... ....388–92; disc. 392
Experimental Study of Magnetic Materials for Use in Ultrahigh-Temperature Electronic Transformers, An. Harms..................181–4; disc. 184
Expressions Involving Complete Elliptic Integrals, The Numerical Evaluation of. Grover........496–502

F

Facility, A Precision, Guarded Resistance Measuring. Wyeth, Higley, Shirk............471–5; disc. 475
Factors Affecting the Use of Over-the-Horizon Links in Telecommunication Networks. Parry.....485–96
Failure in Semiconductor Rectifiers, Forward Current Surge. Gentry................746–50; disc. 754
Failure-Rate Studies on Silicon Rectifiers. Bechtold, Hanks..................................49–56
Fast-Response Full-Wave Magnetic Amplifier, A. Lynn..................................37–41
Fast-Response Magnetic Amplifiers, Volt-Second Transfer Efficiency in. Pula:
  Part I..................................861–7
Fault Protection for Semiconductor Rectifiers, The Current Limiting Fuse as. Gutzwiller............. ....751–4; disc. 754
Feedback in Magnetic Amplifiers, On. Finzi, Suozzi:
  Part I. Single Feedbacks.......1019–30; disc. 1030
Feedback Factor of Self-Balancing Magnetic Amplifiers, The Effective. Geyger..................66–9
Feedback Systems, Binary Communication. Harris, Hauptschein, Morgan, Schwartz..........960–9
Feedback Systems, Binary Symmetric Decision. Harris, Morgan..................................436–43
Feedwater Control System for the Experimental Boiling Water Reactor, An Electronic. Morse, Brey..... 679–83
Ferroresonant Circuit, The. Kelly....843–8; disc. 1061
Ferroresonant Series Circuit Containing a Square-Loop Reactor, Behavior of the. Dennard.......903–11
Field Strength in Oersteds Near Large Rectangular Conductors, Magnetic. Miller.......726–7; disc. 727
Filter Design Applied to Transistorized Carrier System, Insertion Loss. Winkler................519–24
Finite Difference Stability, An Electric Circuit Theory Approach to. Karplus................210–13
Finite Differences to Nonuniformly Spaced Variables, A Generalization of the Calculus of. Kron...539–44
Fins, Thermal Impedance of Cooling. Diebold, Luft... ....739–45; disc. 745
First Cost of a Cable Carrier System, Determination of the Least. Krzyczkowski.......778–82; disc. 782
Five-Channel Telegraphic Tape, Error-Checking for. Barbeau..................................190–3
Flashover Voltage, The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet. (Committee Report)..................350–7; disc. 357
Flux-Gate Magnetometers, Self-Balancing. Geyger... 213–16
Ford Nuclear Reactor, The Design of a Pile Oscillator for the. Stone, Kerr..................1003–06
Formula Studies, Digital Calculation of Network Functions Used in Loss. Ferguson, Long, Rindt....... ....647–52; disc. 652
Forward Current Surge Failure in Semiconductor Rectifiers. Gentry..................746–50; disc. 754
Four-Wire Y 3-Phase System, Watt-Hour Meter for Two Phases of a. McQuarrie.........152–5; disc. 155
Frequency-Converting Telephone Carrier Repeater for Military Use, A. Goltsos, Johnston, Anderson.... 432–6
Frequency Error Correction in Carrier Systems, A New Method of. Stachiewicz................175–80
Frequency Measuring Circuit Accurate to 0.1 Per Cent, An Analog. Mitchell...........983–5; disc. 985
Frequency Modulation Digital Subset for Data Transmission Over Telephone Lines, A. Weber..867–72
Frequency Over-the Horizon Equipment, Wide-Band Ultrahigh-. Felsenheld, Havstad, Jatlow, Levine, Pollack..................................86–93
Full-Wave Magnetic Amplifier, A Fast-Response. Lynn..................................37–41
Function Generator, A Constant-Amplitude Random. Hellwarth..................................443–52
Function Generators Using Short-Time Memory Devices and Nonlinear Elements, The Design of. Revay, Ford..................................143–52
Function Tables in Digital Control Computers. Shubert..................................316–19
Functions in Magnetic-Amplifier Analysis, Core-Reset. Feth:
  Part I. A Core Reset Function....503–09; disc. 509
  Part II. Determination of Amplifier Performance... 510–19
Functions Used in Loss Formula Studies, Digital Calculation of Network. Ferguson, Long, Rindt....... ....647–52; disc. 652
Fundamental Aspects of Ground Resistance Measurements, Some of the. Curdts......760–7; disc. 767
Fuse as Fault Protection for Semiconductor Rectifiers, The Current-Limiting. Gutzwiller............. ....751–4; disc. 754

G

Gaging Plating Thickness, A System for. Myers, Waidelich..................................770–4
General Approach for Obtaining Transient Response by the Use of a Digital Computer, A. Lego, Sze..... 1031–6
Generalization of the Calculus of Finite Differences to Nonuniformly Spaced Variables, A. Kron.539–44
Generating Station, Electrical Features of Indian Point Nuclear Electric. Reimers................879–84
Generator, A Constant-Amplitude Random Function. Hellwarth..................................443–52
Generator, Analysis of Nonresonant Charging Circuit for Magnetic Pulse. Weinberg............271–7
Generator Instrumentation for Nuclear Power Plant with Analog Computer Vertification of Dynamic Action, Steam-. Lynch, Waite............690–6
Generator Reserve Requirement Study, The Use of a Digital Computer in a. Brown..............82–5
Generators, Spectrum Analysis of Random Noise. Polimerou..................................672–9
Generators Using Short-Time Memory Devices and Nonlinear Elements, The Design of Function. Revay, Ford..................................143–52
Ground Measurement, The Technique and Instrumentation of Low-Impedance. Duke, Smith...767–70
Ground Resistance Measurements, Some of the Fundamental Aspects of. Curdts.......760–7; disc. 767
Guarded Resistance Measuring Facility, A Precision. Wyeth, Higley, Shirk............471–5; disc. 475

H

Half-Wave Magnetic Amplifier, Triggering in the Reversible. McMurtrie, Ward..........598–605
Harmonic Magnetic Modulators, Design Criteria for Low-Level Second-. Kletsky............1013–19
Havana Radio System and Its Integration Into the Telephone Networks, The Miami-. Stiles, Hollins, Fruhner, Siddall..................................94–6
High Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive. Boyajian.545–50
High-Efficiency Push-Pull Magnetic Amplifiers with Transistors as Switched Rectifiers. Milnes....... ....327–30; disc. 330

High-Voltage 3-Phase Transformer, Measurement of Voltage Resulting From Single-Phase Switching of a. Dunkle, MacKenzie.........292–4; disc. 294
Horizon Equipment, Wide-Band Ultrahigh-Frequency Over-the-. Felsenheld, Havstad, Jatlow, LeVine, Pollack..........86–93
Horizon Links in Telecommunication Networks, Factors Affecting the Use of Over-the-. Parry..485–96
Hygrometer, Electric. Meiklejohn............302–05

I

Impedance and Return Loss Performance of Telephone Plant in Metropolitan Areas. Bogan..........257–61; disc. 261
Impedance of Cooling Fins, Thermal. Diebold, Luft..........739–45; disc. 745
Impedance Repeaters on Long Rural Telephone Lines, The Application of Negative-. Uthlaut.....230–4
Impedance Telephone Repeaters, Transistorized Negative-. Dimmer..........305–11
Improved 6-Channel Time Division Multiplex for Submarine Cable Telegraphy, An. Wilder....217–22
Impulse Voltage Dividers, Transient Response of. Fisher..........411–19; disc. 419
Impulse Waves with an Isolated Screen Room Installation, Measurements of Steep-Front. Miller, Wittibschlager..........262–70; disc. 270
Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint, Dial Switching of Toll Circuits in. Petrie..........481–5
Indian Point Nuclear Electric Generating Station, Electrical Features of. Reimers..........879–84
Indicator With Elpak Translation Logic, An Electroluminescent Digital. Sack..........113–18
Inductance, Elimination of Magnetic Amplifier Control Circuit. Chandler, Downing.....550–6; disc. 556
Induction-Motor Drive, Transistor-Oscillator. Card..........531–5
Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage, The. (Committee Report)..........350–7; disc. 357
Information, Bits of. Zamanakos..........197–201
Information Theory to Digital Transmission Systems, Recent Progress in. Filipowsky..........848–55
Ink Recorders, Magnetic-Amplifier-Operated. Geyger..........457–71
Insertion Loss Filter Design Applied to Transistorized Carrier System. Winkler..........519–24
Installation, Measurements of Steep-Front Impulse Waves with an Isolated Screen Room. Miller, Wittibschlager..........262–70; disc. 270
Instrumentation for a Nuclear Power Plant with Analog Computer Verification of Dynamic Action, Steam-Generator. Lynch, Waite..........690–6
Instrumentation for High Performance Pressurized Water Reactor Power Plants, Main Coolant. Noyes..........683–90
Instrumentation, Modular Concepts in Reactor Control. Corkrell, Ryan, Magee..........696–9
Instrumentation of Low-Impedance Ground Measurement, The Technique and. Duke, Smith...767–70
(Instruments Electric) Hygrometer. Meiklejohn..........302–05
(Instruments) New Printing Demand Meter. Terry..........80–2
Instruments, On the Application of Zener Diodes to Expanded Scale. Corson..........535–9
Integral, The Numerical Evaluation of Expressions Involving Complete Elliptic. Grover.....496–502
Integrated Mobile Microwave Communications System for the Massachusetts Turnpike, An. Lowe..........703–06; disc. 706
Interference to Microwave Communication Services, Radar. Campbell..........717–22; disc. 722
Interoffice Trunking Facilities, Optimal Utilization and Extension of. Kalaba, Juncosa..........1998–1002; disc. 1002
Intertoll Trunks—Transmission Considerations, Nationwide Switching of. Wallman....223–9; disc. 229
Introducing Young Engineers to the Appreciation of Magnetic Amplifier Problems. Finzi......119–26
Ionosphere, Measurement of the Scattering Matrix with an Intervening. Brysk..........611–12
Iron-Cored Reactor Oscillatory Circuits, Subharmonics in. Mostafa, El-Karaksi..........104–13
Isolated Screen Room Installation, Measurements of Steep-Front Impulse Waves with an. Miller, Wittibschlager..........262–70; disc. 270

K

Key to the Continuing Control Renaissance, The Controlled Rectifier. Harnden....1006–12; disc. 1012

L

Laboratory Tool, The Operational Amplifier as a. Pfeiffer..........947–51; disc. 951
Linear Power Amplifiers Using Dynistors or Trinistors. Hierholzer..........892–8
Linearization of Dead Space. Jones..........836–8
Links to Telecommunication Networks, Factors Affecting the Use of Over-the-Horizon. Parry......485–96
Load Characteristics, Evaluation of Single Phase Metering Practices Based on Meter Performance and. Vanderleck..........727–38; disc. 738
Logic, An Electroluminescent Digital Indicator with Elpak Translation. Sack..........113–18
Logic for Applying Topological Methods to Electric Networks. Byerly, Long, King..657–65; disc. 665
Logical Networks, Matrix Analysis of. Schubert.10–13
Loops on Magnetic Amplifiers, Reactions of Bistable B-H. Morgan..........586–90
Loss Filter Design Applied to Transistorized Carrier System, Insertion. Winkler..........519–24
Loss Formula Studies, Digital Calculation of Network Functions Used in. Ferguson, Long, Rindt..........647–52; disc. 652
Loss Performance of Telephone Plant in Metropolitan Areas, Impedance and Return. Bogan..........257–61; disc. 261
Losses on Loaded Trunk Cables and Methods Taken to Improve Them, Study of Return. Speake..........201–04; disc. 204
Low Impedance Ground Measurement, The Technique and Instrumentation of. Duke, Smith.....767–70
Low-Level Second Haronic Magnetic Modulators, Design Criteria for. Kletsky..........1013–19

M

Magnetic Amplifier, A Fast-Response Full-Wave. Lynn..........37–41
Magnetic-Amplifier Analysis, Core-Reset Functions in. Feth:
  Part I. A Core Reset Function....503–09; disc. 509
  Part II. Determination of Amplifier Performance..........510–19
Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive, High-Efficiency Push-Pull. Boyajian..........545–50
Magnetic Amplifier Bibliography—1951–1956. (Committee Report)..........613–27
Magnetic Amplifier Bibliography, 1957. (Committee Report)..........1051–7
Magnetic-Amplifier Commutating and Pulse-Width Encoding Circuit, A. Lucke..........884–92
Magnetic Amplifier Control Circuit Inductance, Elimination of. Chandler, Downing...550–6; disc. 556
Magnetic Amplifier Controlled Switched Transistors, Reversible-Polarity D-C Power Amplifier Using. Schmitz, Bernstein..........1058–60; disc. 1060
Magnetic Amplifier Cores, Proposed Standards for Core Test Methods for Toroidal. (Committee Report)..........524–31
Magnetic Amplifier Design. Anderson:
  Part I. Core..........160–7
  Part II. The Windings..........167–75
Magnetic Amplifier Design, A Comprehensive Study of. Hubbard, Bishop:
  Part I..........562–72
  Part II..........572–80; disc. 580
Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor, A Power Method of Measuring the Torque-Producing Output of a. Reisman, Bowman..........584–6
Magnetic-Amplifier-Operated Ink Recorders. Geyger..........457–71
Magnetic Amplifier Problems, Introducing Young Engineers to the Appreciation of. Finzi........119–26
Magnetic Amplifier, Triggering in the Reversible Half-Wave. McMurtrie, Ward..........598–605
(Magnetic Amplifiers) A New Control Amplifier Using a Saturable Current Transformer and a Switching Transistor. Morgan..........557–62; disc. 562
Magnetic Amplifiers, A-C Controlled. Lehtonen, Cronauer..........476–80
(Magnetic Amplifiers) Behavior of the Ferroresonant Series Circuit Containing a Square-Loop Reactor. Dennard..........903–11
(Magnetic Amplifiers) Block-Diagram Method of Analysis Applied to the Saturable Reactor. Hubbard..........57–64; disc. 65
Magnetic Amplifiers, On Feedback in. Finzi, Suozzi:
  Part I. Single Feedbacks.......1019–30 disc. 1030
Magnetic Amplifiers, Proposed Standard Terms and Definitions for. (Committee Report)..........429–31; disc. 431
Magnetic Amplifiers, Reactions of Bistable B-H Loops on. Morgan..........586–90
Magnetic Amplifiers, The Effective Feedback Factor of Self-Balancing. Geyger..........66–9
(Magnetic Amplifiers) The Transactor: A Self-Saturated Transformer. Rosenstein:
  Part I..........129–41; disc. 141
(Magnetic Amplifiers) Transistor-Oscillator Induction-Motor Drive. Card..........531–5
Magnetic Amplifiers Utilizing Square-Loop Core Material and Square-Wave Supply Voltage, Characteristics of. Britten..........590–7; disc. 597
Magnetic Amplifiers, Volt-Second Transfer Efficiency in Fast-Response. Pula:
  Part I..........861–7
Magnetic Amplifiers with Transistors as Switched Rectifiers, High-Efficiency Push-Pull. Milnes..........327–30; disc. 330
Magnetic Drum Components for High Storage Density. Hong..........667–72
Magnetic-Drum Storage System Considered for Use as a Common Sender in Nation-Wide Dialing, A. May..........5–10
Magnetic Field Strength in Oersteds Near Large Rectangular Conductors. Miller.....726–7; disc. 727
Magnetic Materials for Use in Ultrahigh-Temperature Electronic Transformers, An Experimental Study of. Harms..........181–4; disc. 184
Magnetic Modulator, A Triple-Action Positioning Controller Using a. Bullen..........785–7
Magnetic Modulators, Design Criteria for Low-Level Second-Harmonic. Kletsky..........1013–19
Magnetic Pulse Generator, Analysis of Nonresonant Charging Circuit for. Weinberg..........271–7
Magnetic Saturable-Core Timing Device. Lowrance..........393–7
Magnetics, Applications of Nonlinear. Storm..........380–8; disc. 388
Magnetometers, Self-Balancing Flux-Gate. Geyger..........213–16
Main Coolant Instrumentation for High-Performance Pressurized Water Reactor Power Plants. Noyes..........683–90
Massachusetts Turnpike, An Integrated Mobile Microwave Communications System for the. Lowe..........703–06; disc. 706
Mathematical Techniques, Some Practical. Carter..........295–302
Matrices to Circuit Theory, Certain Applications of. Pipes..........251–6
Matrix Analysis of Logical Networks. Shubert...10–13
Matrix Programming of Electronic Analog Computers. Horn, Honnell..........420–6; disc. 426
Matrix with an Intervening Ionosphere, Measurement of the Scattering. Brysk..........611–12
Measurement Errors Introduced by Coaxial Cables, Surge. Park..........343–7; disc. 347
Measurement of the Scattering Matrix with an Intervening Ionosphere. Brysk..........611–12
Measurement of the Time Constant and Response Time of Thermal Converts, The Definition and. Hermach..........277–82; disc. 282
Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase Transformer. Dunkle, MacKenzie..........292–4; disc. 294
Measurement, The Technique and Instrumentation of Low-Impedance Ground. Duke, Smith..767–70
Measurements of Abnormal Corona, Some. Penney, Hewitt..........319–27
Measurements of Steep-Front Impulse Waves with an Isolated Screen Room Installation. Miller, Wittibschlager..........262–70: disc. 270
Measurements, Some of the Fundamental Aspects of Ground Resistance. Curdts......760–7; disc. 767
Measuring Circuit Accurate to 0.1 Per Cent, An Analog Frequency. Mitchell..........983–5; disc. 985
Measuring Facility, A Precision, Guarded Resistance. Wyeth, Higley, Shirk..........471–5; disc. 475
Mechanization of Toll Switching at Providence, R.I. Anderson, Thornton..........954–60
Memory Devices and Nonlinear Elements, The Design of Function Generators Using Short-Time. Revay, Ford..........143–52
(Metallic Rectifiers) Failure-Rate Studies on Silicon Rectifiers. Bechtold, Hanks..........49–56
Meter for Two Phases of a 4-Wire Y 3-Phase System, Watt-Hour. McQuarrie..........152–5; disc. 155
Meter, New Printing Demand. Terry..........80–2
Metering Practices Based on Meter Performance and Load Characteristics, Evaluation of Single Phase. Vanderleck..........727–38; disc. 738
Method of Analysis Applied to the Saturable Reactor, Block-Diagram. Hubbard........57–64; disc. 65
Method of Frequency Error Correction in Carrier Systems, A New. Stachiewicz..........175–80
Methods of Code Construction, Diagrammatic. Clark..........817–23
Methods Taken to Improve Them, Study of Return Losses on Loaded Trunk Cables and. Speake..........201–04; disc. 204
Metropolitan Areas, Impedance and Return Loss Performance of Telephone Plant in. Bogan..........257–61; disc. 261
Maimi-Havana Radio System and Its Integration into the Telephone Networks, The. Stiles, Hollins, Fruhner, Siddall..........94–6
Microwave Antenna System, A New Broad-Band. Friis, May..........97–100
Microwave Channel on the Trans-Canada TD-2 System, The Simultaneous Transmission of Television and Telephone Multiplex Over a Single. Curtis, Strahlendorf, Wade..........185–90
Microwave Communication Services, Radar Interference to. Campbell........717–22; dis. 722

Microwave Communications System for the Massachusetts Turnpike, An Integrated Mobile. Lowe ....703–06; disc. 706

Microwave Network, The Expansion of the Pacific Coast. Kuck........898–903

Military Use, A Frequency-Converting Telephone Carrier Repeater for. Goltsos, Johnston, Anderson........432–6

Minimization of Components in Electronic Switching Beatson........283–91; disc. 291

MITE Teleprinter, The. Howard........311–15

Mobile Microwave Communications System for the Massachusetts Turnpike, An Integrated. Lowe ....703–06; disc. 706

Mobile Telephone System, Automatic. Dimmer..723–6

Moderated Reactors for Central Station Power, Organic-Parkins, Weisner........985–93

Modular Concepts in Reactor Control Instrumentation. Cockrell, Ryan, Magee........696–9

Modulator, A Triple-Action Positioning Controller Using a Magnetic. Bullen........785–7

Modulator, The Diode Reactance. Montgomery 680–3

Modulators, Design Criteria for Low-Level Second-Harmonic Magnetic. Kletsky........1013–19

Motor Drive, High-Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt. Boyajian 545–50

Motor Drive, Transistor-Oscillator Induction-. Card 531–5

Multiplex for Submarine Cable Telegraphy, An Improved 6-Channel Time Division. Wilder..217–22

Multiplex Over a Single Microwave Channel On the Trans-Canada TD-2 System, The Simultaneous Transmission of Television and Telephone. Curtis, Strahlendorf, Wade........185–9

N

Nation-Wide Dialing, A Magnetic-Drum Storage System Considered for Use as a Common Sender in. May........5–10

Nationwide Switching of Intertoll Trunks—Transmission Considerations. Wallman......223–9; disc. 229

Negative-Impedance Repeaters on Long Rural Telephone Lines, The Application of. Uthlaut...230-4

Negative-Impedance Telephone Repeaters, Transistorized. Dimmer........305–11

Network Functions Used in Loss Formula Studies, Digital Calculation of. Ferguson, Long, Rindt ....647–52; disc. 652

Network, The Expansion of the Pacific Coast Microwave. Kuck........898–903

Networks, Factors Affecting the Use of Over-the-Horizon Links in Telecommunication. Parry...485–96

Networks, Logic for Applying Topological Methods to Electric. Byerly, Long, King...657–65; disc. 665

Networks, Matrix Analysis of Logical. Schubert..10–13

Networks, The Analysis of Redundancy. Moskowitz 627–32

Networks, The Miami-Havana Radio System and Its Integration into the Telephone. Stiles, Hollins, Fruhner, Siddall........94–6

New Control Amplifier Using a Saturable Current Transformer and a Switching Transistor, A. Morgan ....557–62; disc. 562

New Printing Demand Meter. Terry........80–2

New Small Crossbar Telephone System for Private Branch Exchanges, A. Abbott........911–18

Noise Generators, Spectrum Analysis of. Polimerou 672–9

Noise Signals, Performance of Class B Audio Amplifiers with Random. Usher........939–43

Nonlinear Elements, The Design of Function Generators Using Short Time Memory Devices and. Revay, Ford........143–52

Nonlinear Magnetics, Applications of. Storm........ ....380–8; disc. 388

Nonresonant Charging Circuit for Magnetic Pulse Generator, Analysis of. Weinberg........271–7

Nuclear Electric Generating Station, Electrical Features of Indian Point. Reimers........879–84

Nuclear Power Plant, Analog Computer Study of the Transient Performance of a Dual Cycle Boiling Water Reactor. Leiby........17–25

Nuclear Power Plant with Analog Computer Verification of Dynamic Action, Steam Generator Instrumentation for a. Lynch, Waite........690–6

Nuclear Reactor, The Design of a Pile Oscillator for the Ford. Stone, Kerr........1003–06

(Nucleonic and Radiation Instrumentation) An Electronic Feedwater Control System for the Experimental Boiling Water Reactor. Morse, Brey 679–8

(Nucleonic and Radiation Instrumentation) Main Coolant Instrumentation for High-Performance Pressurized Water Reactor Power Plants. Noyes 683–90

(Nucleonic and Radiation Instrumentation) Modular Concepts in Reactor Control Instrumentation. Cockrell, Ryan, Magee........696–9

(Nucleonics) Organic-Moderated Reactors for Central Station Power. Perkins, Weisner........985–93

Numercial Evaluation of Expressions Involving Complete Elliptic Integrals, The. Grover.....496–502

O

(Oak Ridge) Variation of the Trip Point in the ORNL-Type Safety System. Cockrell, Ricker....699–703

Obtaining Transient Response by the Use of a Digital Computer, A General Approach for. Lego, Sze 1031–6

Oersteds Near Large Rectangular Conductors, Magnetic Field Strength in. Miller......726–7; disc. 727

OJ Carrier Telephone System, Type. Coy, Kelsey, Nicholls........810–17

ON/K Carrier System, The. Masters........806–10

Operation Studies, Digital Computer Aids in Power Pool McIntyre, Blake, Clubb........652–7; disc. 657

Operational Amplifier as a Laboratory Tool, The. Pfeiffer........947–51; disc. 951

Operations Research in Communications. Harris, Hauptschein, Schwartz........633–7; disc. 637

Optimal Utilization and Extension of Interoffice Trunking Facilities. Kalaba, Juncosa........ ....998–1002; disc. 1002

Optimum Block Length for Data Transmission with Error Checking. Wood........855–61

Organic-Moderated Reactors for Central Station Power. Parkins, Weisner........985–93

ORNL-Type Safety System, Variation of the Trip Point in the. Cockrell, Ricker........699–703

Oscillator, An Amplitude-Stabilized Bridged-T. Enslein. 75–9

Oscillator for the Ford Nuclear Reactor, The Design of a Pile. Stone, Kerr........1003–06

Oscillator Induction-Motor Drive, Transistor-. Card 531–5

Oscillators, Transistor Phase-Locked. Edwards, Golubjatnikov, Brady........1043–51

Oscillatory Circuits, Subharmonics in Iron-Cored Reactor. Mostafa, El-Karaksi........104–13

Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor, A Power Method of Measuring the Torque-Producing. Reisman, Bowman. 584–6

Over-the-Horizon Links in Telecommunication Networks, Factors Affecting the Use of. Parry..485–96

Oxide Masked Diffused Structures, Silicon Controlled Rectifiers from. Aldrich, Holonyak........952–4

P

Pacific Coast Microwave Network, The Expansion of the Kuck........898–903

(Paper Tape) Error-Checking for 5-Channel Telegraphic Tape. Barbeau........190–3

Parts Selection. The Economics of Component. Fuegel 13–16

Performance Objectives in a Telephone Communication System, A Basis for Transmission. MacAdam ....205–08; disc. 208

Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant, Analog Computer Study of the Transient. Leiby........17–25

Performance of Class B Audio Amplifiers with Random Noise Signals. Usher........939–43

Performance of Telephone Plant in Metropolitan Areas, Impedance and Return Loss. Bogan........ ....257–61; disc. 261

Performance of the Primary Coolant System in a Water Reactor, Digital Calculation of Transient. Lewis 1036–43

Phase-Locked Oscillators, Transistor. Edwards, Golubjatnikov, Brady........1043–51

Philosophy of Applying Digital Computers to the Design of Electric Apparatus. Abetti, Cuthbertson Williams........367–77; disc. 377

Photographic Electronic Recorder, Digital and Pictorial. MacPherson, Sonderby........194–6

Pictorial Photographic Electronic Recorder, Digital and. McPherson, Sonderby........194–6

Picture Communication, Statistical Encoding for Test and. Michel........33–6

Pile Oscillator for the Ford Nuclear Reactor, The Design of a. Stone, Kerr........1003–6

Plan 55–A, Automatic Telegraph Switching System. Vernam........239–47

Plating Thickness, A System for Gaging. Myers, Waidelich........770-4

Polarity D-C Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors, Reversible-. Schmitz, Bernstein........1058–60; disc. 1060

Pole-Mounted Equipment, Stand-by Power for. Jamieson........774–8

Positioning Controller Using Magnetic Modulator, A Triple-Action. Bullen........785-7

Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors, Reversible-Polarity D-C. Schmitz, Bernstein........1058–60; disc. 1060

Power Amplifiers Using Dynistors or Trinistors, Linear. Hierholzer........892–8

Power for Pole-Mounted Equipment, Stand-by. Jamieson........774–8

Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor, A. Reisman, Bowman 584–6

Power, Organic-Moderated Reactors for Central Station. Parkins, Weisner........985–93

Power Plant with Analog Computer Verification of Dynamic Action, Steam Generator Instrumentation for a Nuclear. Lynch, Waite........690–6

Power Plants, Main Coolant Instrumentation for High-Performance Pressurized Water Reactor. Noyes 683–90

Power Pool Operation Studies, Digital Computer Aids in. McIntyre, Blake, Clubb........652–7; disc. 657

Practical Mathematical Techniques, Some. Carter 295–302

Practice and the Economics of Applying Digital Computers to Engineering Problems, The. Abetti, Williams........331–41; disc. 341

Practices Based on Meter Performance and Load Characteristics, Evaluation of Single-Phase Metering Vanderleck........727–38; disc. 738

Preamplifier, A Stable Direct-Coupled Transistor Servo. DeSautels........943–7

Precipitation Rate Upon 60-Cycle Wet Flashover Voltage, The Influence of Water Resistivity and. Committee Report)........350–7; disc. 357

Precipitation Rectifiers, Automatic Control of Electrical. Van Hoesen, White, Hall........126–8

Precision, Guarded Resistance Measuring Facility, A. Wyeth, Higley, Shirk........471–5; disc. 475

Pressurized Water Reactor Power Plants, Main Coolant Instrumentation for High-Performance. Noyes 683–90

Primary Coolant System in a Water Reactor, Digital Calculation of Transient Performance of the. Lewis........1036–43

Printing Demand Meter, New. Terry........80–2

Private Branch Exchange, A New Small Crossbar Telephone System for. Abbott........911–18

Problems, Introducing Young Engineers to the Appreciation of Magnetic Amplifier. Finzi........119–26

Problems, The Practice and the Economics of Applying Digital Computers to Engineering. Abetti, Williams........331–41; disc. 341

Programming of Electronic Analog Computers, Matrix. Horn, Honnell........420–6; disc. 426

Progress in Applying Information Theory to Digital Transmission Systems, Recent. Filipowsky.848–55

Proposed Standard Terms and Definitions for Magnetic Amplifiers. (Committee Report)........ ....429–31; disc. 431

Proposed Standards for Core Test Methods for Toroidal Magnetic Amplifier Cores. (Committee Report).. 524–31

Protection for Semiconductor Rectifiers, The Current-Limiting Fuse as Fault. Gutzwiller........ ....751–4; disc. 754

Providence, R.I., Mechanization of Toll Switching at. Anderson, Thornton........954–60

Public Air–Ground Telephone Service Trial. Augustus 926–30

Pulse Generator, Analysis of Nonresonant Charging Circuit for Magnetic. Weinberg........271–7

Pulse-Width Encoding Circuit, A Magnetic-Amplifier Commutating and. Lucke........884–92

Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive, High-Efficiency. Boyajian 545–50

Push-Pull Magnetic Amplifiers With Transistors as Switched Rectifiers, High-Efficiency. Milnes.. ....327–30; disc. 330

Q

Quadruplexer Allowing the Simultaneous Transmission of Two Complete Television Stations Using a Common Antenna, A. MacKimmie........787–91

R

Radar Interference to Microwave Communication Services. Campbell........717–22; disc. 722

Radio Attenuation at 11 Kmc and Some Implications Affecting Relay System Engineering. Hathaway, Evans........930–8; disc. 1061

(Radio Communication) A New Broad-Band Microwave Antenna System. Friis, May........97–100

(Radio Communication) An Integrated Mobile Microwave Communications System for the Massachusetts Turnpike. Lowe........703–06; disc. 706

(Radio Communication) Dial Telephone Service for Smith Island, an Isolated Community in Chesapeake Bay. Littleton........923–6

(Radio Communication) Factors Affecting the Use of Over-the-Horizon Links in Telecommunication Networks. Parry........485–96

(Radio Communication) Public Air-Ground Telephone Service Trial. Augustus....926–30
(Radio Communication) Radar Interference to Microwave Communication Services. Campbell....717–22; disc. 722
(Radio Communication) The Expansion of the Pacific Coast Microwave Network. Kuck....898–903
(Radio Communication) The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System. Curtis, Strahlendorf, Wade.185–90
(Radio Communication) Tropospheric-Scatter System Design. Yeh....707–16
(Radio Communication) Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. Felsenheld, Havstad, Jatlow, LeVine, Pollack....86–93
Radio System and Its Integration into the Telephone Networks, The Miami-Havana. Stiles, Hollins, Fruhner, Siddall....94–6
Radio System for the United States Corps of Engineers, Willamette Valley. Pedersen, Marihart...401–06
Radio System for Toll Telephone Service, A 6,000-Megacycle. Kebby, Culbertson....969–79
(Rain Tests) The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage. (Committee Report)...350–7; disc. 357
Random Function Generator, A Constant-Amplitude. Hellwarth....443–52
Random Noise Generators, Spectrum Analysis of. Polimerou....672–9
Random Noise Signals, Performance of Class B Audio Amplifiers with. Usher....939–43
Random Selectors, Spinning Disk. Roberts....755–60
Rate Studies on Silicon Rectifiers, Failure-. Bechtold, Hanks....49–56
Reactance Modulator, The Diode. Montgomery.980–3
Reactions of Bistable B-H Loops on Magnetic Amplifiers. Morgan....586–90
Reactor, An Electronic Feedwater Control System for the Experimental Boiling Water. Morse, Brey....679–83
Reactor, Behavior of the Ferroresonant Series Circuit Containing a Square-Loop. Dennard.....903–11
Reactor, Block-Diagram Method of Analysis Applied to the Saturable. Hubbard....57–64; disc. 65
Reactor Control Instrumentation, Modular Concepts in. Cockrell, Ryan, Magee....696–9
Reactor, Digital Calculation of Transient Performance of the Primary Coolant System in a Water. Lewis....1036–43
Reactor Nuclear Power Plant, Analog Computer Study of the Transient Performance of a Dual Cycle Boiling Water. Leiby....17–25
Reactor Oscillatory Circuits, Subharmonics in Iron-Cored. Mostafa, El-Karaksi....104–13
Reactor Power Plants, Main Coolant Instrumentation for High Performance Pressurized Water. Noyes....683–90
Reactors for Central Station Power, Organic-Moderated. Parkins, Weisner....985–93
Reactors for Semiconductor Rectifiers, Current-Balancing. Dortort....452–6; disc. 456
Recent Progress in Applying Information Theory to Digital Transmission Systems. Filipowsky.848–55
Recorder, Digital and Pictorial Photographic Electronic. McPherson, Sonderby....194–6
Recorders, Magnetic-Amplifier-Operated Ink. Geyger....457–71
(Recording and Controlling Instruments) A Triple-Action Positioning Controller Using a Magnetic Modulator. Bullen....785–7
Rectangular Conductors, Magnetic Field Strength in Oersteds Near Large. Miller.....726–7; disc. 727
Rectifier: Key to the Continuing Control Renaissance, The Controlled. Harnden....1006–12; disc. 1012
Rectifiers, Automatic Control of Electrical Precipitation. VanHoesen, White, Hall....126–8
Rectifiers, Current-Balancing Reactors for Semiconductor. Dortort....452–6; disc. 456
Rectifiers, Failure-Rate Studies on Silicon. Bechtold, Hanks....49–56
Rectifiers from Oxide Masked Diffused Structures, Silicon-Controlled. Aldrich, Holonyak.....952–4
Rectifiers, Forward Current Surge Failure in Semiconductor. Gentry....746–50; disc. 754
Rectifiers, High-Efficiency Push-Pull Magnetic Amplifiers with Transistors as Switched. Milnes....327–30; disc. 330
Rectifiers, The Current Limiting Fuse as Fault Protection for Semiconductor. Gutzwiller....751–4; disc. 754
Redundancy Networks, The Analysis of. Moskowitz....627–32
Relay Scanning Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements. Kay....248–50
Relay System Engineering, Radio Attenuation at 11 Kmc and Some Implications Affecting. Hathaway, Evans....930–8; disc. 1061
Reliable Stochastic Sequential Switching Circuits. Mullin....606–11
Renaissance, The Controlled Rectifier: Key to the Continuing Control. Harnden....1006–12; disc. 1012
Repeater for Military Use, A Frequency-Converting Telephone Carrier. Goltsos, Johnston, Anderson....432–6
Repeater for Use with the 45BN Cable Carrier System, A Transistorized. Babin, Fish....41–9
Repeaters on Long Rural Telephone Lines, The Application of Negative-Impedance. Uthlaut......230–4
Repeaters, Transistorized Negative-Impedance Telephone. Dimmer....305–11
Requirement Study, The Use of a Digital Computer in a Generator Reserve. Brown....82–5
Research in Communications, Operations. Harris, Hauptschein, Schwartz....633–7; disc. 637
Reserve Requirement Study, The Use of a Digital Computer in a Generator. Brown....82–5
Reset Functions in Magnetic-Amplifier Analysis, Core-. Feth:
  Part I. A Core Reset Function....503–09; disc. 509
  Part II. Determination of Amplifier Performance....510–19
Resistance Measuring Facility, A Precision, Guarded. Wyeth, Higley, Shirk....471–5; disc. 475
Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage, The Influence of Water. (Committee Report)....350–7; disc. 357
Response of Impulse Voltage Dividers, Transient. Fisher....411–19; disc. 419
Response Time of Thermal Converters, The Definition and Measurement of the Time Constant and. Hermach....277–82; disc. 282
Return Losses on Loaded Trunk Cables and Methods Taken to Improve Them, Study of. Speake....201–04; disc. 204
Reversible Half-Wave Magnetic Amplifier, Triggering in the. McMurtrie, Ward....598–605
Reversible-Polarity D-C Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors. Schmitz, Bernstein....1058–60; disc. 1060
Rural Telephone Lines, The Application of Negative Impedance Repeaters on Long. Uthlaut....230–4

S

Safety System, Variation of the Trip Point in the ORNL-Type. Cockrell, Ricker....699–703
SAGE Data System Considerations. Enticknap, Schuster....824–32
SAGE Data Systems, Communication Channels for. James....838–43
SAGE Data Terminals. Soffel, Spack....872–9
Saturable-Core Timing Device, Magnetic. Lowrance....393–7
Saturable Current Transformer and a Switching Transistor, A New Control Amplifier Using a. Morgan....557–62; disc. 562
Saturable Reactor, Block-Diagram Method of Analysis Applied to the. Hubbard....57–64; disc. 65
Scale Instruments, On the Application of Zener Diodes to Expanded. Corson....535–9
Scanning Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements, Relay. Kay....248–50
Scatter System Design, Tropospheric-. Yeh....707–16
Scattering Matrix with an Intervening Ionosphere, Measurement of the. Brysk....611–12
School Teaching, Television Systems for In-. Kraus....406–11
Screen Room Installation, Measurements of Steep-Front Impulse Waves With an Isolated. Miller, Wittibschlager....262–70; disc. 270
Second Harmonic Magnetic Modulators, Design Criteria for Low-Level. Kletsky....1013–19
Selection, The Economics of Component Parts. Fuegel....13–16
Selectors, Spinning Disk Random. Roberts....755–60
Self-Balancing Flux-Gate Magnetometers. Geyger....213–16
Self-Balancing Magnetic Amplifiers, The Effective Feedback Factor of. Geyger....66–9
Self-Saturated Transformer, The Transactor: A. Rosenstein:
  Part I....129–41; disc. 141
(Semiconductor Metallic Rectifiers) Linear Power Amplifiers Using Dynistors or Trinistors. Hierholzer....892–8
(Semiconductor Metallic Rectifiers) The Controlled Rectifier: Key to the Continuing Control Renaissance. Harnden....1006–12; disc. 1012
(Semiconductor Metallic Rectifiers) Theory of Transient Build-up in Avalanche Transistors. Shockley, Gibbons....993–8
Semiconductor Rectifiers, Current-Balancing Reactors for. Dortort....452–6; disc. 456
Semiconductor Rectifiers, Forward Current Surge Failure in. Gentry....746–50; disc. 754
Semiconductor Rectifiers, The Current Limiting Fuse as Fault Protection for. Gutzwiller....751–4; disc. 754
Serial-Type Digital Data Systems, An Analog Computer Study of the Transient Behavior and Stability Characteristics of. Elgerd......358–66; disc. 366
Series Circuit Containing a Square-Loop Reactor, Behavior of the Ferroresonant. Dennard.....903–11
Servo Motor, A Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase. Reisman, Bowman....584–6
Servo Preamplifier, A Stable Direct-Coupled Transistor. DeSautels....943–7
Set, Data Transmission Testing. Boughtwood, Christie....101–04
Shifting Counters. Eldert, Gray, Gurk, Rubinoff..70–4
Short-Time Memory Devices and Nonlinear Elements, The Design of Function Generators Using. Revay, Ford....143–52
Shunt Motor Drive, High Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C. Boyajian..545–50
Signals, Performance of Class B Audio Amplifiers with Random Noise. Usher....939–43
Silicon-Controlled Rectifiers from Oxide Masked Diffused Structures. Aldrich, Holonyak.......952–4
Silicon Rectifiers, Failure-Rate Studies on. Bechtold, Hanks....49–56
Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System, The. Curtis, Strahlendorf, Wade....185–90
Simultaneous Transmission of Two Complete Television Stations Using a Common Antenna, A Quadruplexer Allowing the. MacKimmie........787–91
Single Phase Metering Practices Based on Meter Performance and Load Characteristics, Evaluation of. Vanderleck....727–38; disc. 738
Single-Phase Switching of a High-Voltage 3-Phase Transformer, Measurement of Voltage Resulting From. Dunkle, MacKenzie.....292–4; disc. 294
Six-Channel Time Division Multiplex for Submarine Cable Telegraphy, An Improved. Wilder..217–22
60-Cycle Wet Flashover Voltage, The Influence of Water Resistivity and Precipitation Rate Upon. (Committee Report)....350–7; disc. 357
6,000-Megacycle Radio System for Toll Telephone Service, A. Kebby, Culbertson....969–79
Smith Island, an Isolated Community in Chesapeake Bay, Dial Telephone Service for. Littleton..923–6
Some of the Fundamental Aspects of Ground Resistance Measurements. Curdts....760–7; disc. 767
Some Practical Mathematical Techniques. Carter....295–302
Space, Linearization of Dead. Jones....836–8
Spaced Variables, A Generalization of the Calculus of Finite Differences to Nonuniformly. Kron.539–44
Spectrum Analysis of Random Noise Generators. Polimerou....672–9
Spinning Disk Random Selectors. Roberts....755–60
Square-Loop Core Material and Square-Wave Supply Voltage, Characteristics of Magnetic Amplifiers Utilizing. Britten....590–7; disc. 597
Square-Loop Reactor, Behavior of the Ferroresonant Series Circuit Containing a. Dennard.....903–11
Stability, An Electric Circuit Theory Approach to Finite Difference. Karplus....210–13
Stability Characteristics of Serial-Type Digital Data Systems, An Analog Computer Study of the Transient Behavior and. Elgerd.....358–66; disc. 366
Stabilized Bridged-T Oscillator, An Amplitude-. Enslein....75–9
Stabilized D-C Differential Transistor Amplifier, A. Depian, Smith....157–9
Stable Direct-Coupled Transistor Servo Preamplifier, A. DeSautels....943–7
Standard Terms and Definitions for Magnetic Amplifiers, Proposed. (Committee Report)....429–31; disc. 431
Standards for Core Test Methods for Toroidal Magnetic Amplifier Cores, Proposed. (Committee Report)....524–31
Stand-by Power for Pole-Mounted Equipment. Jamieson....774–8
Statistical Encoding for Text and Picture Communication. Michel....33–6
Steam-Generator Instrumentation for a Nuclear Power Plant with Analog Computer Verification of Dynamic Action. Lynch, Waite....690–6
Steep-Front Impulse Waves with an Isolated Screen Room Installation, Measurements of. Miller, Wittibschlager....262–70; disc. 270
Stochastic Sequential Switching Circuits, Reliable. Mullin....606–11
Storage Density, Magnetic Drum Components for High. Hong....667–72
Storage System Considered for Use as a Common Sender in Nation-Wide Dialing, A Magnetic-Drum. May....5–10
Structures, Silicon Controlled Rectifiers from Oxide Masked Diffused. Aldrich, Holonyak......952–4
Studies on Silicon Rectifiers, Failure-Rate. Bechtold, Hanks....49–56
Study of a Binary Code, An Experimental. Peterson....388–92; disc. 392
Study of Magnetic-Amplifier Design, A Comprehensive. Hubbard, Bishop:
  Part I....562–72
  Part II....572–80; disc. 580

Vol,
Pt. 1
195

Study of Magnetic Materials for Use in Ultrahigh-Temperature Electronic Transformers, An Experimental. Harms....181–4; disc. 184
Study of Return Losses on Loaded Trunk Cables and Methods Taken to Improve Them. Speake....201–04; disc. 204
Study of the Transient Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant, Analog Computer. Leiby....17–25
Study, The Use of a Digital Computer in a Generator Reserve Requirement. Brown....82–5
Subharmonics in Iron-Cored Reactor Oscillatory Circuits. Mostafa, El-Karaksi....104–13
Submarine Cable Telegraphy, An Improved 6-Channel Time Division Multiplex for. Wilder....217–22
Subset for Data Transmission Over Telephone Lines, A Frequency Modulation Digital. Weber....867–72
Surge Failure in Semiconductor Rectifiers, Forward Current. Gentry....746–50; disc. 754
Surge Measurement Errors Introduced by Coaxial Cables. Park....343–7; disc. 347
Switched Rectifiers, High-Efficiency Push-Pull Magnetic Amplifiers with Transistors as. Milnes....327–30; disc. 330
Switched Transistors, Reversible-Polarity D-C Power Amplifier Using Magnetic-Amplifier Controlled. Schmitz, Bernstein....1058–60; disc. 1060
Switching at Providence, R.I., Mechanization of Toll. Anderson, Thornton....954–60
Switching Circuits, Minimization of Components in Electronic. Beatson....283–91; disc. 291
Switching Circuits, Reliable Stochastic Sequential. Mullin....606–11
Switching of a High-Voltage 3-Phase Transformer, Measurement of Voltage Resulting From Single-Phase Dunkle, MacKenzie....292–4; disc. 294
Switching of Interoll Trunks—Transmission Considerations, Nationwide. Wallman....223–9; disc. 229
Switching of Toll Circuits, in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint, Dial. Petrie....481–5
Switching System Plan 55-A, Automatic Telegraph. Vernam....239–47
Switching Transistor, A New Control Amplifier Using a Saturable Current Transformer and a. Morgan....557–62; disc. 562
Symmetric Channel, On Coding for the Binary. Fontaine, Peterson....638–46; disc. 646
Symmetric Decision Feedback Systems, Binary. Harris, Morgan....436–43
Synchronized Clocks for Data Transmission. Edson, Flavin, Perry....832–6
Synthesis by Means of Computers, Direct System. Clymer....798–806
System Considerations, SAGE Data. Enticknap, Schuster....824–32
System Design, Tropospheric-Scatter. Yeh....707–16
System Engineering, Radio Attenuation at 11 Kmc and Some Implications Affecting Relay. Hathaway, Evans....930–8; disc. 1061
System for Gaging Plating Thickness, A. Myers, Waidelich....770–4
System Synthesis by Means of Computers, Direct. Clymer....798–806

T

Tables in Digital Control Computers, Function. Shubert....316–19
Tape, Error-Checking for 5-Channel Telegraphic. Barbeau....190–3
TD-2 System, The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada. Curtis, Strahlendorf, Wade....185–90
Teaching, Television Systems for In-School. Kraus....406–11
Technique and Instrumentation of Low-Impedance Ground Measurement, The. Duke, Smith....767–70
Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements, Relay Scanning Design. Kay....248–50
Techniques, Accurate Determination of Capacitance by Conjoining Analytical and Analog. Horgan, Pesavento....397–400
Techniques, Some Practical Mathematical. Carter....295–302
Telecommunication Networks, Factors Affecting the Use of Over-the-Horizon Links in. Parry....485–96
Telegraph Switching System Plan 55-A, Automatic. Vernam....239–47
(Telegraph Systems) Airborne Teletypewriter AN/AGC-1. Michals....234–9
(Telegraph Systems) Data Transmission Testing Set. Boughtwood, Christie....101–04
(Telegraph Systems) The MITE Teleprinter. Howard....311–15
Telegraphic Tape, Error-Checking for 5-Channel. Barbeau....190–3
Telegraphy, An Improved 6-Channel Time Division Multiplex for Submarine Cable. Wilder....217–22
(Telemetering) Diagrammatic Methods of Code Construction. Clark....817–23
(Telephone) A Transistorized Repeater for Use with the 45BN Cable Carrier System. Babin, Fish....41–9
Telephone Carrier Repeater for Military Use, A Frequency-Converting. Goltsos, Johnston, Anderson....432–6
Telephone Communication System, A Basis for Transmission Performance Objectives in a. MacAdam....205–08; disc. 208
Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing, Standpoint, Dial Switching of Toll Circuits in Independent. Petrie....481–5
Telephone Lines, A Frequency Modulation Digital Subset for Data Transmission Over. Weber....867–72
Telephone Lines, The Application of Negative-Impedance Repeaters on Long Rural. Uthlaut....230–4
Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System, The Simultaneous Transmission of Television and. Curtis, Strahlendorf, Wade....185–90
Telephone Networks, The Miami-Havana Radio System and its Integration the. Stiles, Hollins, Fruhner, Siddall....94–6
Telephone Plant In Metropolitan Areas, Impedance and Return Loss Performance of. Bogan....257–61; disc. 261
Telephone Repeaters, Transistorized Negative-Impedance. Dimmer....305–11
Telephone Service, A 6,000-Megacycle Radio System for Toll. Kebby, Culbertson....969–79
Telephone Service for Smith Island, an Isolated Community in Chesapeake Bay, Dial. Littleton....923–6
Telephone Service Trial, Public Air-Ground. Augustus....926–30
Telephone System, Automatic Mobile. Dimmer....723–6
Telephone System for Private Branch Exchanges, A New Small Crossbar. Abbott....911–18
Telephone System, Type OJ Carrier. Coy, Kelsey, Nicholls....810–17
Teleprinter, The MITE. Howard....311–15
Teletypewriter AN/AGC-1, Airborne. Michals....234–9
Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System, The Simultaneous Transmission of. Curtis, Strahlendorf, Wade....185–90
Television Broadcasting, Video Testing Techniques in. Ste-Marie....26–33
Television Stations Using a Common Antenna, A Quadruplexer Allowing the Simultaneous Transmission of Two Complete. MacKimmie....787–91
Television Systems for In-School Teaching. Kraus....406–11
Temperature Electronic Transformers, An Experimental Study of Magnetic Materials for Use in Ultrahigh-. Harms....181–4; disc. 184
Terminals, SAGE Data. Soffel, Spack....872–9
Terms and Definitions for Magnetic Amplifiers, Proposed Standard. (Committee Report)....429–31; disc. 431
Test Methods for Toroidal Magnetic Amplifier Cores, Proposed Standards for Core. (Committee Report)....524–31
Testing Set, Data Transmission. Boughtwood, Christie....101–04
Testing Techniques in Television Broadcasting, Video. Ste-Marie....26–33
Text and Picture Communication, Statistical Encoding for. Michel....33–6
Theory Approach to Finite Difference Stability, An Electric Ciruit. Karplus....210–13
Theory, Certain Applications of Matrices to Circuit. Pipes....251–6
Theory of Transient Build-up in Avalanche Transistors. Shockley, Gibbons....993–8
Theory to Digital Transmission Systems, Recent Progress in Applying Information. Filipowsky....848–55
Thermal Converters, The Definition and Measurement of the Time Constant and Response Time of. Hermach....277–82; disc. 282
Thermal Impedance of Cooling Fins. Diebold, Luft....739–45; disc. 745
Thickness, A System for Gaging Plating. Myers, Waidelich....770–4
Three-Phase System, Watt-Hour Meter for Two-Phases of a 4-Wire Y. McQuarrie....152–5; disc. 155
Three-Phase Transformer, Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage. Dunkle, MacKenzie....292–4; disc. 294
Time Constant and Response Time of Thermal Converters, The Definition and Measurement of the. Hermach....277–82; disc. 282
Time Division Multiplex for Submarine Cable Telegraphy, An Improved 6-Channel. Wilder....217–22
Timing Device, Magnetic Saturable-Core. Lowrance....393–7
Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint, Dial Switching of. Petrie....481–5
Toll Switching at Providence, R.I., Mechanization of. Anderson, Thornton....954–60
Toll Telephone Service, A 6,000-Megacycle Radio System for. Kebby, Culbertson....969–79
Tool, The Operational Amplifier as a Laboratory. Pfeiffer....947–51; disc. 951
Topological Methods to Electric Networks, Logic for Applying. Byerly, Long, King....657–65; disc. 665
Toroidal Magnetic Amplifier Cores, Proposed Standards for Core Test Methods for. (Committee Report)....524–31
Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor, A Power Method of Measuring the. Reisman, Bowman....584–6
Transactor: A Self-Saturated Transformer, The. Rosenstein:
  Part I....129–41; disc. 141
Transducer Measurements, Relay Scanning Design Technique Generates High Accuracy and Speed in Analog-to-Digital. Kay....248–50
Transfer Efficiency in Fast-Response Magnetic Amplifiers, Volt-Second. Pula:
  Part I....861–7
Transformer and a Switching Transistor, A New Control Amplifier Using a Saturable Current. Morgan....557–62; disc. 562
Transformer, Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase. Dunkle, MacKenzie....292–4; disc. 294
Transformer, The Transactor: A Self-Saturated. Rosenstein:
  Part I....129–41; disc. 141
Transformers, An Experimental Study of Magnetic Materials for Use in Ultra-high Temperature Electronic. Harms....181–4; disc. 184
Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems, An Analog Computer Study of the. Elgerd....358–66; disc. 366
Transient Build-up in Avalanche Transistors, Theory of. Shockley, Gibbons....993–8
Transient Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant, Analog Computer Study of the. Leiby....17–25
Transient Response by the Use of a Digital Computer, A General Approach for Obtaining. Lego, Sze....1031–6
Transient Response of Impulse Voltage Dividers. Fisher....411–19; disc. 419
Transistor, A New Control Amplifier Using a Saturable Current Transformer and a Switching. Morgan....557–62; disc. 562
Transistor Amplifier, A Stabilized D-C Differential. Depian, Smith....157–9
Transistor-Oscillator Induction-Motor Drive. Card....531–5
Transient Performance of the Primary Coolant System in a Water Reactor, Digital Calculation of. Lewis....1036–43
Transistor Phase-Locked Oscillators. Edwards, Golubjatnikov, Brady....1043–51
Transistor Servo Preamplifier, A Stable Direct Coupled. DeSautels....943–7
Transistorized Carrier System, Insertion Loss Filter Design Applied to. Winkler....519–24
Transistorized Compandor, A. Perkins, Perreault, Perkins....791–7
Transistorized Negative-Impedance Telephone Repeaters. Dimmer....305–11
Transistorized Repeater for Use with the 45BN Cable Carrier System, A. Babin, Fish....41–9
Transistors as Switched Rectifiers, High-Efficiency Push-Pull Magnetic Amplifiers with. Milnes....327–30; disc. 330
Transistors, Reversible-Polarity D-C Power Amplifier Using Magnetic Amplifier Controlled Switched. Schmitz, Bernstein....1058–60, disc. 1060
Transistors, Theory of Transient Build-up in Avalanche. Shockley, Gibbons....993–8
Translation Logic, An Electroluminescent Digital Indicator with Elpak. Sack....113–18
Transmission Considerations from the Direct Distance Dialing Standpoint, Dial Switching of Toll Circuits in Independent Telephone Company Areas:. Petrie....481–5
Transmission Considerations, Nationwide Switching of Intertoll Trunks—. Wallman....223–9; disc. 229
Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System, The Simultaneous. Curtis, Strahlendorf, Wade....185–90
Transmission of Two Complete Television Stations Using a Common Antenna, A Quadruplexer Allowing the Simultaneous. MacKimmie....787–91
Transmission Over Telephone Lines, A Frequency Modulation Digital Subset for Data. Weber....867–72
Transmission Performance Objectives in a Telephone Communication System, A Basis for. MacAdam....205–08; disc. 208
Transmission, Synchronized Clocks for Data. Edson, Flavin, Perry....832–6
Transmission Systems, Recent Progress in Applying Information Theory to Digital. Filipowsky....848–55

Transmission Testing Set, Data. Boughtwood, Christie 101–04
Transmission with Error Checking, Optimum Block Length for Data. Wood........ 855–61
Triggering in the Reversible Half-Wave Magnetic Amplifier. McMurtrie, Ward........ 598–605
Trinistors, Linear Power Amplifiers Using Dynistors or. Hierholzer........ 892–8
Trip Point in the ORNL-Type Safety System, Variation of the. Cockrell, Ricker........ 699–703
Triple-Action Positioning Controller Using a Magnetic Modulator, A. Bullen........ 785–7
Tropospheric-Scatter System Design. Yeh........ 707–16
Trunk Cables and Methods Taken to Improve Them, Study of Return Losses on Loaded. Speake........ 201–04; disc. 204
Trunking Facilities, Optimal Utilization and Extension of Interoffice. Kalaba, Juncosa........ 998–1002; disc. 1002
Trunks—Transmission Considerations, Nationwide Switching of Intertoll. Wallman. 223–9; disc. 229
Type OJ Carrier Telephone System. Coy, Kelsey, Nicholls........ 810–17

U

Ultrahigh-Frequency Over-the-Horizon Equipment, Wide-Band. Felsenheld, Havstad, Jatlow, Levine, Pollack........ 86–93
Ultrahigh-Temperature Electronic Transformers, An Experimental Study of Magnetic Materials for Use in. Harms........ 181–4; disc. 184
United States Corps of Engineers, Willamette Valley Radio System for the. Pedersen, Marihart. 401–06
Use of a Digital Computer in a Generator Reserve Requirement Study, The. Brown........ 82–5
Utilization and Extension of Interoffice Trunking Facilities, Optimal. Kalaba, Juncosa........ 998–1002; disc. 1002

V

Variables, A Generalization of the Calculus of Finite Differences to Nonuniformly Spaced. Kron........ 539–44
Variation of the Trip Point in the ORNL-Type Safety System. Cockrell, Ricker........ 699–703
Video Testing Techniques in Television Broadcasting. Ste-Marie........ 26–33
Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers. Pula:
  Part I........ 861–7
Voltage, Characteristics of Magnetic Amplifiers Utilizing Square-Loop Core Material and Square-Wave Supply. Britten........ 590–7; disc. 597
Voltage Dividers, Transient Response of Impulse. Fisher........ 411–19; disc. 419
Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase Transformer, Measurement of. Dunkle, MacKenzie........ 292–4; disc. 294
Voltage, The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover. (Committee Report)........ 350–7; disc. 357

W

Water Reactor, Digital Calculation of Transient Performance of the Primary Coolant System in a. Lewis........ 1036–43
Water Reactor Nuclear Power Plant, Analog Computer Study of the Transient Performance of a Dual Cycle Boiling. Leiby........ 17–25
Water Reactor Power Plants, Main Coolant Instrumentation for High Performance Pressurized. Noyes........ 683–90
Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage, The Influence of. (Committee Report)........ 350–7; disc. 357
Watt-Hour Meter for Two Phases of a 4-Wire Y 3-Phase System. McQuarrie........ 152–5; disc. 155
Waves with an Isolated Screen Room Installation, Measurements of Steep-Front Impulse. Miller, Wittibschlager........ 262–70; disc. 270
Wet Flashover Voltage, The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle. (Committee Report)........ 350–7; disc. 357
Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. Felsenheld, Havstad, Jatlow, LeVine, Pollack........ 86–93
Willamette Valley Radio System for the United States Corps of Engineers. Pedersen, Marihart... 401–06
(Wire Communications) A Basis for Transmission Performance Objectives in a Telephone Communication System. MacAdam........ 205–08; disc. 208
(Wire Communications) A Frequency-Converting Telephone Carrier Repeater for Military Use. Goltsos, Johnston, Anderson........ 432–6
(Wire Communications) A New Method of Frequency Error Correction in Carrier Systems. Stachiewicz........ 175–80
(Wire Communications) A Transistorized Compandor. Perkins, Perreault, Perkins........ 791–7
(Wire Communications) A Transistorized Repeater for Use with the 45BN Cable Carrier System. Babin, Fish........ 41–9
(Wire Communications) Determination of the Least First Cost of a Cable-Carrier System. Krzyczkowski........ 778–82; disc. 782
(Wire Communications) Dial Switching of Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint. Petrie........ 481–5
(Wire Communications) Impedance and Return Loss Performance of Telephone Plant in Metropolitan Areas. Bogan........ 257–61; disc. 261
(Wire Communications) Insertion Loss Filter Design Applied to Transistorized Carrier System. Winkler........ 519–24
(Wire Communications) Stand-by Power for Pole-Mounted Equipment. Jamieson........ 774–8
(Wire Communications) Study of Return Losses on Loaded Trunk Cables and Methods Taken to Improve Them. Speake........ 201–04; disc. 204
(Wire Communications) The Application of Negative-Impedance Repeaters on Long Rural Telephone Lines. Uthlaut........ 230–4
(Wire Communications) The ON/K Carrier System. Masters........ 806–10
(Wire Communications) Transistorized Negative-Impedance Telephone Repeaters. Dimmer... 305–11
(Wire Communications) Type OJ Carrier Telephone System. Coy, Kelsey, Nicholls........ 810–17

Y

Y Three-Phase System, Watt-Hour Meter for Two Phases of a 4-Wire. McQuarrie.. 152–5; disc. 155

Z

Zener Diodes to Expanded Scale Instruments, On the Application of. Corson........ 535–9

2. Author Index

Abbott, H. H. A New Small Crossbar Telephone System for Private Branch Exchanges. (58-1243)........ 911–18
Abetti, P. A.; W. J. Cuthbertson, S. B. Williams. Philosophy of Applying Digital Computers to the Design of Electric Apparatus. (58–233)........ 367–77; disc. 379
Abetti, P. A.; S. B. Williams. The Practice and the Economics of Applying Digital Computers to Engineering Problems. (58–162)........ 331–41; disc. 341
Aldrich, R. W.; N. Holonyak, Jr. Silicon-Controlled Rectifiers from Oxide Masked Diffused Structures. (58–1260)........ 952–4
Anderson, C. W.; H. N. Thornton. Mechanization of Toll Switching at Providence, R.I. (58–1246)........ 954–60
Anderson, R. B.; G. Goltsos, J. H. Johnston. A Frequency-Converting Telephone Carrier Repeater for Military Use. (58–215)........ 432–6
Anderson, R. E. Magnetic Amplifier Design:
  Part I. The Core. (57–922)........ 160–7
  Part II. The Windings. (58–39)........ 167–75
Anderson, R. E. Disc........ 431
Augustus, L. M. Public Air-Ground Telephone Service Trial. (58–1293)........ 926–30

B

Babin, V.; R. Fish. A Transistorized Repeater for Use with the 45BN Cable Carrier System. (57–907)........ 41–9
Barbeau, R. A. Error-Checking for 5-Channel Telegraphic Tape. (58–2)........ 190–3
Barrow, Bruce B. Disc........ 400
Beatson, T. J. Minimization of Components in Electronic Switching Circuits. (58–3)........ 283–91; disc. 291
Bechtold, N. F.; C. L. Hanks. Failure-Rate Studies on Silicon Rectifiers. (57–795)........ 49–56
Bellaschi, P. L. Disc........ 295
Bennon, S. Disc........ 294
Bernstein, T. Disc........ 562
Bernstein, T.; K. Mei. Disc........ 1062
Bernstein, T.; N. L. Schmitz. Reversible-Polarity D-C Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors. (58–1177)........ 1058–60; disc. 1060
Bernstein, T.; N. L. Schmitz. Disc........ 330
Bills, G. W. Disc........ 657, 666
Bishop, M. M.; R. M. Hubbard. A Comprehensive Study of Magnetic Amplifier Design:
  Part I. (58–792)........ 562–72
  Part II. (58–793)........ 572–80; disc. 580
Blake, C. W.; H. M. McIntyre, J. S. Clubb. Digital Computer Aids in Power Pool Operation Studies. (58–892)........ 652–7; disc. 657
Bogan, L. B. Impedance and Return Loss Performance of Telephone Plant in Metropolitan Areas. (58–4)........ 257–61
Bonn, N. E. Disc........ 475
Boughtwood, J. E.; T. A. Christie. Data Transmission Testing Set. (58–172)........ 101–04
Bourne, E. A. Disc........ 230
Bourne, H. C., Jr.; D. Nitzan. Disc........ 597
Bowman, E. H.; S. J. Reisman. A Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor. (58–897)........ 584–6
Boyajian, C. L. High Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive. (58–775)........ 545–50
Brady, D. J.; K. A. Edwards, O. Golubjatnikov, Transistor Phase-Locked Oscillators. (58–1276)........ 1043–51
Brey, R. N., Jr.; R. H. Morse. An Electronic Feedwater Control System for the Experimental Boiling Water Reactor. (58–700)........ 679–83
Britten, H. H. Characteristics of Magnetic Amplifiers Utilizing Square-Loop Core Material and Square-Wave Supply Voltage. (58–856). 590–7; disc. 598
Brown, Homer E. The Use of a Digital Computer in a Generator Reserve Requirement Study. (58–134)........ 82–5
Brysk, Henry. Measurement of the Scattering Matrix with an Intervening Ionosphere. (58–532). 611–12
Bullen, C. V. A Triple-Action Positioning Controller Using a Magnetic Modulator. (58–850).... 785–7
Byerly, R. T.; R. W. Long, C. W. King. Logic for Applying Topological Methods to Electric Networks. (58–826)........ 657–65; disc. 667

C

Campbell, R. D. Radar Interference to Microwave Communication Services. (58–817)........ 717–22; disc. 722
Card, W. H. Transistor-Oscillator Induction-Motor Drive. (58–898)........ 531–5
Cantrell, Harry N. Disc........ 378
Carter, Gordon K. Some Practical Mathematical Techniques. (57–194)........ 295–302
Chandler, D. P.; R. W. Downing. Elimination of Magnetic Amplifier Control Circuit Inductance. (58–774)........ 550–6
Chapin, Ned. Disc........ 342, 378
Christie, T. A.; J. E. Boughtwood. Data Transmission Testing Set. (58–172)........ 101–04
Clark, Robert C. Diagrammatic Methods of Code Construction. (58–997)........ 817–23
Clubb, J. S.; H. M. McIntyre, C. W. Blake. Digital Computer Aids in Power Pool Operation Studies. (58–892)........ 652–7; disc. 657
Clymer, A. Ben. Direct System Synthesis by Means of Computers. (58–1003)........ 798–806
Cockrell, J. L.; C. W. Ricker. Variation of the Trip Point in the ORNL-Type Safety System. (58–709)........ 699–703
Cockrell, J. L.; D. F. Ryan, J. H. Magee. Modular Concepts in Reactor Control Instrumentation. (58–708)........ 696–9
Corson, A. J. On the Application of Zener Diodes to Expanded Scale Instruments. (58–934)..... 535–9
Costogue, Ernest N. Disc........ 1030
Cox, James R. Disc........ 951
Coy, J. A.; E. S. Kelsey, B. P. Nicholls. The OJ Carrier Telephone System. (58–1004)........ 810–17
Crawshaw, R. C. Disc........ 209
Cronauer, E. A.; E. W. Lehtonen. A-C Controlled Magnetic Amplifiers. (58–773)........ 476–80
Culbertson, A. F.; M. H. Kebby. A 6,000-Megacycle Radio System for Toll Telephone Service. (58–181)........ 969–79
Curdts, E. B. Some of the Fundamental Aspects of Ground Resistance Measurements. (58–106)........ 760–7; disc. 767
Curtis, H. E.; U. C. P. Strahlendorf, A. J. Wade. The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System. (58–183)........ 185–90
Cuthbertson, W. J.; P. A. Abetti, S. B. Williams. Philosophy of Applying Digital Computers to the Design of Electric Apparatus. (58–233)........ 367–77; disc. 377

D

Dantzig, George B. Disc........ 1002
Decker, R. O. Disc........ 1030
Dellinger, Flay G. Disc........ 204

Dennard, R. H. Behavior of the Ferroresonant Series Circuit Containing a Square-Loop Reactor. (58-1215)......903-11
Depian, Louis; R. E. Smith. A Stabilized D-C Differential Transistor Amplifier. (58-192)....157-9
DeSautels, Albert N. A Stable Direct-Coupled Transistor Servo Preamplifier. (58-1238)......943-7
Diebold, E. J.; W. Luft. Thermal Impedance of Cooling Fins. (58-926)......739-45
Dimmer, R. P. Automatic Mobile Telephone System. (57-1020)......723-6
Dimmer, Robert P. Transistorized Negative-Impedance Telephone Repeaters. (57-1016)......305-11
Disney, V. H. Disc......428
Dortort, I. K. Current-Balancing Reactors for Semiconductor Rectifiers. (58-291)...452-6; disc. 456
Dortort, I. K. Disc......745, 754
Downing, R. W.; D. P. Chandler. Elimination of Magnetic Amplifier Control Circuit Inductance. (58-774)......550-6
Downing, W. C., Jr. Disc......282
Duke, C. A.; L. E. Smith. The Technique and Instrumentation of Low-Impedance Ground Measurement. (58-105)......767-70
Dunkle, W. F.; W. F. MacKenzie. Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase Transformer. (58-35)......292-4; disc. 294

E

Edson, J. O.; M. A. Flavin, A. D. Perry. Synchronized Clocks for Data Transmission. (58-300)....832-6
Edwards, K. A.; O. Golubjatnikov, D. J. Brady. Transistor Phase-Locked Oscillators. (58-1276)......1043-51
Eldert, C.; H. J. Gray, Jr., H. M. Gurk, Morris Rubinoff. Shifting Counters. (57-199)......70-4
Elgerd, Olle I. An Analog Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems. (58-31).358-66
Elias, P. Disc......646
El-Karaksi, M. I.; A. E. Mostafa. Subharmonics in Iron-Cored Reactor Oscillatory Circuits. (57-745)......104-13
Enslein, Kurt. An Amplitude-Stabilized Bridged-T Oscillator. (57-76)......75-9
Enticknap, R. G.; E. F. Schuster. SAGE Data System Considerations. (58-302)......824-32
Evans, H. W.; S. D. Hathaway. Radio Attenuation at 11 Kmc and Some Implications Affecting Relay System Engineering. (58-1236).930-8; disc. 1061

F

Farmanfarma, G.; F. Mullin. Disc......366
Felsenheld, R. A.; H. Havstad, J. L. Jatlow, D. J. LeVine, L. Pollack. Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. (58-179)......86-93
Ferguson, R. W.; R. W. Long, L. J. Rindt. Digital Calculation of Network Functions Used in Loss Formula Studies. (58-890).....647-52; disc. 652
Feth, G. C. Core-Reset Functions in Magnetic-Amplifier Analysis:
  Part I. A Core-Reset Function. (57-95)......503-09; disc. 509
  Part II. Determination of Amplifier Performance. (57-921)......510-19
Filipowsky, R. Recent Progress in Applying Information Theory to Digital Transmission Systems. (58-1157)......848-55
Finzi, L. A. Introducing Young Engineers to the Appreciation of Magnetic Amplifier Problems. (58-177)......119-26
Finzi, L. A., J. J. Suozzi. On Feedback in Magnetic Amplifiers:
  Part I. Single Feedbacks. (58-1231)......1019-30; disc. 1030
Fish, R.; V. Babin. A Transistorized Repeater for Use with the 45BN Cable Carrier System. (57-907)......41-9
Fisher, F. A. Transient Response of Impulse Voltage Dividers. (58-37)......411-19; disc. 420
Fisher, F. A.; J. S. Kreage. Disc......348
Fisher, F. A.; A. F. Rohlfs, J. Kreage. Disc......270
Flavin, M. A.; J. O. Edson, A. D. Perry. Synchronized Clocks for Data Transmission. (58-300)....832-6
Fontaine, A. B.; W. W. Peterson. On Coding for the Binary Symmetric Channel. (58-925)......638-46; disc. 646
Ford, D. J.; A. W. Revay. The Design of Function Generators Using Short-Time Memory Devices and Nonlinear Elements. (58-247)......143-52
Fowler, A. D.; R. A. Gibby. Assessment of Effects of Delay Distortion in Data Systems. (58-1240)......918-23
Frey, H. A. Disc......357
Francke, H. C. Disc......1061
Friis, R. W.; A. S. May. A New Broad-Band Microwave Antenna System. (58-185)......97-100
Fruhner, E. T.; K. P. Stiles, F. G. Hollins, W. D. Siddall. The Miami-Havana Radio System and Its Integration into the Telephone Networks. (58-154)......94-6
Fuegel, E. F. The Economics of Component Parts Selection. (57-678)......13-16

G

Gentry, F. E. Forward Current Surge Failure in Semiconductor Rectifiers. (58-927)..746-50; disc. 755
George, E. E. Disc......652, 665
Geyger, W. A. Magnetic-Amplifier-Operated Ink Recorders. (58-541)......457-71
Geyger, W. A. The Effective Feedback Factor of Self-Balancing Magnetic Amplifiers. (57-767)...66-9
Geyger, William A. Self-Balancing Flux-Gate Magnetometers. (58-96)......213-16
Gibbons, J.; W. Shockley. Theory of Transient Build-up in Avalanche Transistors. (58-1249)......993-8
Gibby, R. A.; A. D. Fowler. Assessment of Effects of Delay Distortion in Data Systems. (58-1240)......918-23
Golds, L. B. S. Disc......738
Goltsos, G.; J. H. Johnston, R. B. Anderson. A Frequency-Converting Telephone Carrier Repeater for Military Use. (58-215)......432-6
Golubjatnikov, O.; K. A. Edwards, D. J. Brady. Transistor Phase-Locked Oscillators. (58-1276)......1043-51
Gray, H. J., Jr.; C. Eldert, H. M. Gurk, Morris Rubinoff. Shifting Counters. (57-199)......70-4
Grosch, H. R. J. Disc......341, 377
Grover, Frederick W. The Numerical Evaluation of Expressions Involving Complete Elliptic Integrals. (58-777)......496-502
Gurk, H. M.; C. Eldert, H. J. Gray, Jr., Morris Rubinoff. Shifting Counters. (57-199)......70-4
Gutzwiller, F. W. The Current Limiting Fuse as Fault Protection for Semiconductor Rectifiers. (58-928)......751-4; disc. 755
Gutzwiller, F. W. Disc......1012

H

Halacsy, A. A. Disc......452
Hall, H. J.; H. E. VanHoesen, H. J. White. Automatic Control of Electrical Precipitation Rectifiers. (58-229)......126-8
Hanks, C. L.; N. F. Bechtold. Failure-Rate Studies on Silicon Rectifiers. (57-795)......49-56
Harms, H. B. An Experimental Study of Magnetic Materials for Use in Ultrahigh-Temperature Electronic Transformers. (58-209)...181-4; disc. 184
Harnden, J. D., Jr. The Controlled Rectifier: Key to the Continuing Control Renaissance. (58-1234)......1006-12
Harris, B.; A. Hauptschein, K. C. Morgan, L. S. Schwartz. Binary Communication Feedback Systems. (58-1233)......960-9
Harris, B.; A. Hauptschein, L. S. Schwartz. Operations Research in Communications. (58-575)....633-7
Harris, Bernard; K. C. Morgan. Binary Symmetric Decision Feedback Systems. (57-1041)....436-43
Hathaway, S. D.; H. W. Evans. Radio Attentuation at 11 Kmc and Some Implications Affecting Relay System Engineering. (58-1236).930-8; disc. 1061
Hauptschein, A.; B. Harris, K. C. Morgan, L. S. Schwartz. Binary Communication Feedback Systems. (58-1233)......960-9
Hauptschein, A.; B. Harris, L. S. Schwartz. Operations Research in Communications. (58-575)....633-7
Havstad, H.; R. A. Felsenheld, J. L. Jatlow, D. J. LeVine, L. Pollack. Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. (58-179)......86-93
Hellwarth, G. A. A Constant-Amplitude Random Function Generator. (58-533)......443-52
Herder, D. J. Disc......379
Hermach, Francis L. The Definition and Measurement of the Time Constant and Response Time of Thermal Converters. (58-11)......277-82; disc. 282
Hewitt, Jack G., Jr.; G. W. Penney. Some Measurements of Abnormal Corona. (57-87)......319-27
Hierholzer, Frank J., Jr. Linear Power Amplifiers Using Dynistors or Trinistors. (58-1206)......892-8
Higley, J. B.; F. H. Wyeth, W. H. Shirk, Jr. A Precision, Guarded Resistance Measuring Facility. (58-556)......471-5; disc. 475
Hodtum, J. B. Disc......295
Hollins, F. G.; K. P. Stiles, E. T. Fruhner, W. D. Siddall. The Miami-Havana Radio System and Its Integration into the Telephone Networks. (58-154)..94-6
Holonyak, N., Jr.; R. W. Aldrich. Silicon-Controlled Rectifiers from Oxide Masked Diffused Structures. (58-1260)......952-4
Hong, Kyonghi. Magnetic Drum Components for High Storage Density. (58-929)......667-72
Honnell, P. M.; R. E. Horn. Matrix Programming of Electronic Analog Computers. (58-66)......420-6; disc. 428
Horgan, J. D.; J. A. Pesavento. Accurate Determination of Capacitance by Conjoining Analytical and Analog Techniques. (58-65)......397-400
Horn, R. E.; P. M. Honnell. Matrix Programming of Electronic Analog Computers. (58-66)......420-6; disc. 428
Howard, Bernard. The MITE Teleprinter. (58-48)......311-15
Howard, J. D., Jr.; J. W. Joyner. Disc......208
Hubbard, R. M. Block-Diagram Method of Analysis Applied to the Saturable Reactor. (57-923)......57-64; disc. 65
Hubbard, R. M. Disc......141, 556
Hubbard, R. M.; M. M. Bishop. A Comprehensive Study of Magnetic Amplifier Design:
  Part I. (58-792)......562-72
  Part II. (58-793)......572-80; disc. 580
Hylten-Cavallius, N. Disc......419

J

James, R. T. Communication Channels for SAGE Data Systems. (58-332)......838-43
Jamieson, D. F. Stand-by Power for Pole-Mounted Equipment. (58-851)......774-8
Jatlow, J. L.; R. A. Felsenheld, H. Havstad, D. J. LeVine, L. Pollack. Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. (58-179)......86-93
Johnson, Wheeler X. A Differentiator for A-C Computers. (57-193)......1-4
Johnston, J. H.; G. Goltsos, R. B. Anderson. A Frequency-Converting Telephone Carrier Repeater for Military Use. (58-215)......432-6
Jones, Sheldon. Linearization of Dead Space. (58-1174)......836-8
Joyner, J. W.; J. D. Howard, Jr. Disc......208
Juncosa, M. L.; R. E. Kalaba. Optimal Utilization and Extension of Interoffice Trunking Facilities. (58-1245)......998-1002; disc. 1003

K

Kalaba, R. E.; M. L. Juncosa. Optimal Utilization and Extension of Interoffice Trunking Facilities. (58-1245)......998-1002; disc. 1003
Karplus, Walter J. An Electric Circuit Theory Approach to Finite Difference Stability. (58-60)......210-13
Kay, Andrew F. Relay Scanning Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements. (57-675)......248-50
Kebby, M. H.; A. F. Culbertson. A 6,000-Megacycle Radio System for Toll Telephone Service. (58-181)......969-79
Kelly, George E., Jr. The Ferroresonant Circuit. (58-1168)......843-8; disc. 1061
Kelsey, E. S.; J. A. Coy, B. P. Nicholls. The OJ Carrier Telephone System. (58-1004)......810-17
Kerr, W.; J. Stone. The Design of a Pile Oscillator for the Ford Nuclear Reactor. (58-1329)......1003-06
Kessler, W. J. Disc......706
King, C. W.; R. T. Byerly, R. W. Long. Logic for Applying Topological Methods to Electric Networks. (58-826)......657-65; disc. 667
Kirchmayer, L. K.; A. J. Wood. Disc......666
Kletsky, E. J. Design Criteria for Low-Level Second-Harmonic Magnetic Modulators. (58-1277)......1013-19
Korn, Edward. Disc......580
Kramer, Max. Disc......727
Kraus, Max H. Television Systems for In-School Teaching. (57-699)......406-11
Kreage, J. S.; F. A. Fisher. Disc......348
Kreage, J.; A. F. Rohlfs, F. A. Fisher. Disc......270
Kron, Gabriel. A Generalization of the Calculus of Finite Differences to Nonuniformly Spaced Variables. (58-787)......539-44
Krzyczkowski, R. Disc......637
Krzyczkowski, Roman. Determination of the Least First Cost of a Cable-Carrier System. (58-905)......778-82; disc. 782
Kuck, R. G. The Expansion of the Pacific Coast Microwave Network. (58-1227)......898-903

L

Lantz, Martin J. Disc......667
Lebens, J. C. Disc......754
Lego, P. E.; T. W. Sze. A General Approach for Obtaining Transient Response by The Use of a Digital Computer. (58-1291)......1031-6
Lehtonen, E. W.; E. A. Cronauer. A-C Controlled Magnetic Amplifiers. (58-773)......476-80
Leiby, David W. Analog Computer Study of the Transient Performance of a Dual Cycle Boiling Water Reactor Nuclear Power Plant. (57-684)......17-25

LeVine, D. J.; R. A. Felsenheld, H. Havstad, J. L. Jatlow, L. Pollack. Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment. (58–179)........86–93
Lewis, D. G. Digital Calculation of Transient Performance of the Primary Coolant System in a Water Reactor. (58–1314)........1036–43
Littleton, M. E. Dial Telephone Service for Smith Island, and Isolated Community in Chesapeake Bay. (58–1292)........923–6
Long, R. W.; R. T. Byerly, C. W. King. Logic for Applying Topological Methods to Electric Networks. (58–826)........657–65; disc. 667
Long, R. W.; R. W. Ferguson, L. J. Rindt. Digital Calculation of Network Functions Used in Loss Formula Studies. (58–890).....647–52; disc. 652
Lord, H. W. Disc........509
Lowe, Ralph F. An Integrated Mobile Microwave Communications System for the Massachusetts Turnpike. (58–815)........703–06; disc. 706
Lowenschuss, Oscar. Disc........291
Lowrance, J. L. Magnetic Saturable-Core Timing Device. (58–120)........393–7
Lucke, W. H. A Magnetic Amplifier Commutating and Pulse-Width Encoding Circuit. (58–1162)..884–92
Luft, Werner; E. J. Diebold. Thermal Impedance of Cooling Fins. (58–926)........739–45
Lynch, E. E.; D. P. Waite. Steam-Generator Instrumentation for a Nuclear Power Plant with Analog Computer Verification of Dynamic Action. (58–703)........690–6
Lynn, G. E. A Fast-Response Full-Wave Magnetic Amplifier. (57–773)........37–41

M

MacAdam, W. K. A Basis for Transmission Performance Objectives in a Telephone Communication System. (58–59)........205–08; disc. 209
MacKenzie, W. F.; W. F. Dunkle. Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase Transformer. (58–35)........292–4; disc. 294
MacKimmie, G. B. A Quadruplexer Allowing the Simultaneous Transmission of Two Complete Television Stations Using a Common Antenna. (58–904)........787–91
Magee, J. H.; J. L. Cockrell, D. F. Ryan. Modular Concepts in Reactor Control Instrumentation. (58–708)........696–9
Marihart, D. J.; C, Pedersen. Willamette Valley Radio System for the United States Corps of Engineers. (57–914)........401–06
Masters, G. T. The ON/K Carrier System. (58–1005)........806–10
May, A. S.; R. W. Friis. A New Broad-Band Microwave Antenna System. (58–185)........97–100
May, H. F. A Magnetic-Drum Storage System Considered for Use as a Common Sender in Nation-Wide Dialing. (56–660)........5–10
McCluskey, E. J., Jr. Disc........291
McCreary, H. J. Disc........388
McIntyre, H. M.; C. W. Blake, J. S. Clubb. Digital Computer Aids in Power Pool Operation Studies. (58–892)........652–7; disc. 657
McMurtrie, D. L.; J. P. Ward. Triggering in the Reversible Half-Wave Magnetic Amplifier. (58–776)........598–605
McPherson, R. G.; I. A. Sonderby. Digital and Pictorial Photographic Electronic Recorder. (58–29)........194–6
McQuarrie, A. M. Watt-Hour Meter for Two Phases of a 4-Wire Y 3-Phase System. (58–145)........152–5; disc. 156
Meahl, Harry R. Disc........722
Mei, K.; T. Bernstein. Disc........1062
Meiklejohn, W. H. Electric Hygrometer. (58–8)........302–05
Michals, Richard A. Airborne Teletypewriter AN/AGC-1. (58–47)........234–9
Michel, W. S. Statistical Encoding for Text and Picture Communication. (57–723)........33–6
Mikkelsen, S. E.; R. P. Thompson. Disc........1002
Miller, C. J., Jr. Disc........358
Miller, C. J., Jr.; J. F. Wittibschlager. Measurements of Steep-Front Impulse Waves with an Isolated Screen Room Installation. (58–18)........262–70; disc. 271
Miller, Charles J.; J. F. Wittibschlager. Disc......347
Miller, J. H. Magnetic Field Strength in Oersteds Near Large Rectangular Conductors. (58–367)........726–7
Miller, John H.; J. Nagy. Disc........282
Milnes, A. G. High-Efficiency Push-Pull Magnetic Amplifiers with Transistors as Switched Rectifiers. (58–77)........327–30; disc. 330
Milnes, A. G. Disc........65
Mitchell, J. An Analog Frequency Measuring Circuit Accurate to 0.1 Per Cent. (58–1275)........983–5; disc. 985
Montgomery, G. Franklin. The Diode Reactance Modulator. (58–1326)........980–3
Morey, C. G. Disc........156
Morgan, K. C.; B. Harris. Binary Symmetric Decision Feedback Systems. (57–1041)........436–43
Morgan, K. C.; B. Harris, A. Hauptschein, L. S. Schwartz. Binary Communication Feedback Systems. (58–1233)........960–9
Morgan, R. E. Reactions of Bistable B-H Loops on Magnetic Amplifiers. (57–776)........586–90
Morgan, Ray E. A New Control Amplifier Using a Saturable Current Transformer and a Switching Transistor. (58–858)........557–62; disc. 562
Morse, R. H.; R. N. Brey, Jr. An Electronic Feedwater Control System for the Experimental Boiling Water Reactor. (58–700)........679–83
Moskowitz, Fred. The Analysis of Redundancy Networks. (58–538)........627–32
Mostafa, Abd El-Samie; M. I. El-Karaksi. Subharmonics in Iron-Cored Reactor Oscillatory Circuits. (57–745)........104–13
Mullin, Albert A. Reliable Stochastic Sequential Switching Circuits. (58–854)........606–11
Mullin, F.; G. Farmanfarma. Disc........366
Myers, R. G.; D. L. Waidelich. A System for Gaging Plating Thickness. (58–933)........770–4

N

Nagy, John; J. H. Miller. Disc........282
Nelson, E. E.; G. Rezos. Disc........783
Nicholls, B. P.; J. A. Coy, E. S. Kelsey. The OJ Carrier Telephone System. (58–1004)........810–17
Nitzan, D.; H. C. Bourne, Jr. Disc........597
Noyes, L. C. Main Coolant Instrumentation for High-Performance Pressurized Water Reactor Power Plants. (58–702)........683–90

P

Park, J. H. Surge Measurement Errors Introduced by Coaxial Cables. (58–110)........343–7; disc. 350
Park, J. H. Disc........270, 419
Parkins, W. E.; E. F. Weisner. Organic-Moderated Reactors for Central Station Power. (58–1272)........985–93
Parry, Charles A. Factors Affecting the Use of Over-the-Horizon Links in Telecommunication Networks. (58–383)........485–96
Pasnak, M. Disc........184
Paynter, H. M. Disc........952
Paynter, Henry M.; D. H. Sheingold. Disc........426
Pedersen, C.; D. J. Marihart. Willamette Valley Radio System for the United States Corps of Engineers. (57–914)........401–06
Penney, Gaylord W.; J. G. Hewitt, Jr. Some Measurements of Abnormal Corona. (57–87)......319–27
Perkins, A. F.; J. C. Perkins, Jr., D. A. Perreault. A Transistorized Compandor. (58–906)......791–7
Perkins, J. C., Jr.; D. A. Perreault, A. F. Perkins. A Transistorized Compandor. (58–906)......791–7
Perreault, D. A.; J. C. Perkins, Jr., A. F. Perkins. A Transistorized Compandor. (58–906)......791–7
Perry, A. D.; J. O. Edson, M. A. Flavin. Synchronized Clocks for Data Transmission. (58–300)....832–6
Pesavento, J. A.; J. D. Horgan. Accurate Determination of Capacitance by Conjoining Analytical and Analog Techniques. (58–65)........397–400
Peterson, W. W. An Experimental Study of a Binary Code. (58–38)........388–92
Peterson, W. W.; A. B. Fontaine. On Coding for the Binary Symmetric Channel. (58–925)........638–46; disc. 646
Petrie, J. N. Dial Switching of Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint. (58–386)........481–5; disc. 783
Pfeiffer, Paul E. The Operational Amplifier as a Laboratory Tool. (58–1198)........947–51; disc. 952
Pipes, Louis A. Certain Applications of Matrices to Circuit Theory. (58–52)........251–6
Pipes, Louis A. Disc........426
Polimerou, L. G. Spectrum Analysis of Random Noise Generators. (57–701)........672–9
Pollack, L.; R. A. Felsenheld, H. Havstad, J. L. Jatlow, D. J. LeVine. Wide-Band Ultrahigh Frequency Over-the-Horizon Equipment. (58–179)........86–93
Pula, T. J. Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers:
  Part I. (58–1193)........861–7

R

Reimers, T. D. Electrical Features of Indian Point Nuclear Electric Generating Station. (58–1194)........879–84
Reisman, S. J.; E. H. Bowman. A Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor. (58–897)........584–6
Rekoff, M. G., Jr. Disc........1060
Revay, A. W.; D. J. Ford. The Design of Function Generators Using Short-Time Memory Devices and Nonlinear Elements. (58–247)........143–52
Rezos, G. Disc........782
Rezos, George; E. E. Nelson. Disc........783
Ricker, C. W.; J. L. Cockrell. Variation of the Trip Point in the ORNL-Type Safety System. (58–709)........699–703
Rideout, Vincent C. Disc........426
Rindt, L. J.; R. W. Ferguson, R. W. Long. Digital Calculation of Network Functions Used in Loss Formula Studies. (58–890).....647–52; dis. 652
Roberts, G. A. Spinning Disk Random Selectors. (58–534)........755–60
Robertson, George B. M. Disc........155
Rohlfs, A. F. Disc........349
Rohlfs, A. F.; J. Kresge, F. A. Fisher. Disc........270
Rose, Arthur F. Disc........229, 261, 783
Rosenstein, Allen B. The Transactor: A Self-Saturated Transformer:
  Part I. (58–178)........129–41; disc. 142
Rubinoff, Morris; C. Eldert, H. J. Gray, Jr., H. M. Gurk. Shifting Counters. (57–199)........70–4
Ryan, D. F.; J. L. Cockrell, J. H. Magee. Modular Concepts in Reactor Control Instrumentation. (58–708)........696–9

S

Sack, E. A. An Electroluminescent Digital Indicator with Elpak Translation Logic. (58–150)........113–18
Ste-Marie, A. Video Testing Techniques in Television Broadcasting. (57–666)........26–33
Schmitz, N. L.; T. Bernstein. Reversible-Polarity D-C Power Amplifier Using Magnetic-Amplifier Controlled Switched Transistors. (58–1177)........1058–60; disc. 1060
Schmitz, N. L.; T. Bernstein. Disc........330
Schubert, E. J. Function Tables in Digital Control Computers. (58–231)........316–19
Schubert, E. J. Matrix Analysis of Logical Networks. (57–673)........10–13
Schubert, Ernest J. Disc........666
Schuster, E. F.; R. G. Enticknap. SAGE Data System Considerations. (58–302)........824–32
Schwartz, L. S.; B. Harris, A. Hauptschein. Operations Research in Communications. (58–575).....633–7
Schwartz, L. S.; B. Harris, A. Hauptschein, K. C. Morgan. Binary Communication Feedback Systems. (58–1233)........960–9
Shea, R. F. Disc........400
Shedd, Paul C. Disc........706
Sheingold, Daniel H.; H. M. Paynter. Disc........426
Shirk, W. H., Jr.; F. H. Wyeth, J. B. Higley. A Precision, Guarded Resistance Measuring Facility. (58–556)........471–5; disc. 475
Shockley, W.; J. Gibbons. Theory of Transient Build-up in Avalanche Transistors. (58–1249)........993–8
Siddall, W. D.; K. P. Stiles, F. G. Hollins, E. T. Fruhner. The Miami-Havana Radio System and Its Integration into the Telephone Networks. (58–154)........94–6
Smith, L. E.; C. A. Duke. The Technique and Instrumentation of Low-Impedance Ground Measurement. (58–105)........767–70
Smith, Richard E.; L. Depian. A Stabilized D-C Differential Transistor Amplifier. (58–192)....157–9
Soffel, R. O.; E. G. Spack. SAGE Data Terminals. (58–407)........872–9
Sonderby, I. A.; R. G. McPherson. Digital and Pictorial Photographic Electronic Recorder. (58–29)........194–6
Spack, E. G.; R. O. Soffel. SAGE Data Terminals. (58–407)........872–9
Speake, G. H. Study of Return Losses on Loaded Trunk Cables and Methods Taken to Improve Them. (58–214)........201–04; disc. 204
Spitzer, C. F. Disc........1061
Stachiewicz, B. R. A New Method of Frequency Error Correction in Carrier Systems. (58–239)...175–80
Stiles, K. P.; F. G. Hollins, E. T. Fruhner, W. D. Siddall. The Miami-Havana Radio System and Its Integration into the Telephone Networks. (58–154)........94–6
Stone, J.; W. Kerr. The Design of a Pile Oscillator for the Ford Nuclear Reactor. (58–1329)....1003–06
Storm, H. F. Disc........431
Storm, Herbert F. Applications of Nonlinear Magnetics. (58–180)........380–8
Strahlendorf, U. C. P.; H. E. Curtis, A. J. Wade. The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System. (58-183)........185–90
Suozzi, J. J.; L. A. Finzi. On Feedback in Magnetic Amplifiers:
  Part I. Single Feedbacks. (58–1231)........1019–30; disc. 1030
Sze, T. W.; P. E. Lego. A General Approach for Obtaining Transient Response by the Use of a Digital Computer. (58–1291)........1031–6

T

Tagg, G. F. Disc. 767
Terry, P. V. New Printing Demand Meter. (58–146) 80–2
Thompson, P. M. Disc. 985
Thompson, R. P.; S. E. Mikkelsen. Disc. 1002
Thornton, H. N.; C. W. Anderson. Mechanization of Toll Switching at Providence, R.I. (58–1246) 954–60

U

Usher, Theron, Jr. Performance of Class B Audio Amplifiers with Random Noise Signals. (58–1237) 939–43
Uthlaut, H. T., Jr. The Application of Negative-Impedance Repeaters on Long Rural Telephone Lines. (57–391) 230–4

V

Van Hoesen, H. E., H. J. White, H. J. Hall. Automatic Control of Electrical Precipitation Rectifiers. (58–229) 126–8
Vanderleck, J. M. Evaluation of Single-Phase Metering Practices Based on Meter Performance and Load Characteristics. (58–425) .... 727–38; disc. 738
Veinott, C. G. Disc. 341, 378
Vernam, G. S. Automatic Telegraph Switching System Plan 55-A. (58–45) 239–47

W

Wade, A. J.; H. E. Curtis, U. C. P. Strahlendorf. The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System. (58–183) 185–90
Waidelich, D. L.; R. G. Myers. A System for Gaging Plating Thickness. (58–933) 770–4
Waite, D. P.; E. E. Lynch. Steam-Generator Instrumentation for a Nuclear Power Plant with Analog Computer Vertification of Dynamic Action. (58–703) 690–6
Wallman, O. F. Nationwide Switching of Intertoll Trunks—Transmission Considerations. (58–5) 223–9
Ward, J. P.; D. L. McMurtrie. Triggering in the Reversible Half-Wave Magnetic Amplifier. (58–776) 598–605
Weber, Laurance A. A Frequency Modulation Digital Subset for Data Transmission Over Telephone Lines. (58–1204) 867–72
Weinberg, N. L. Analysis of Nonresonant Charging Circuit for Magnetic Pulse Generator. (57–963) 271–7
Weisner, E. F.; W. E. Parkins. Organic-Moderated Reactors for Central Station Power. (58–1272) 985–93
Whitaker, T. N. Disc. 951
White, H. J.; H. E. Van Hoesen, H. J. Hall. Automatic Control of Electrical Precipitation Rectifiers. (58–229) 126–8
Wilder, H. F. An Improved 6-Channel Time Division Multiplex for Submarine Cable Telegraphy. (58–44) 217–22
Wilhelm, H. T. Disc. 475
Wilkinson, R. I. Disc. 1002
Williams, S. B.; P. A. Abetti. The Practice and the Economics of Applying Digital Computers to Engineering Problems. (58–162) .... 331–41; disc. 341
Williams, S. B.; P. A. Abetti, W. J. Cuthbertson. Philosophy of Applying Digital Computers to the Design of Electric Apparatus. (58–233) .... 367–77; disc. 377
Winkler, T. Insertion Loss Filter Design Applied to Transistorized Carrier System. (57–1070) 519–24
Wittibschlager, J. F.; C. J. Miller, Jr. Measurements of Steep-Front Impulse Waves with an Isolated Screen Room Installation. (58–18) .... 262–70; disc. 271
Wittibschlager, John F.; C. J. Miller, Jr. Disc. 347
Wood, A. J.; L. K. Kirchmayer. Disc. 666
Wood, F. B. Optimum Block Length for Data Transmission with Error Checking. (58–1181) 855–61
Wyeth, F. H.; J. B. Higley, W. H. Shirk, Jr. A Precision, Guarded Resistance Measuring Facility. (58–556) .... 471–5; disc. 475

Y

Yeh, Leang P. Tropospheric-Scatter System Design. (58–816) 707–16

Z

Zamanakos, A. S. Bits of Information. (58–34) 197–201

Vo
Pt.
19

A Differentiator for A-C Computers

WHEELER X. JOHNSON
ASSOCIATE MEMBER AIEE

THIS PAPER describes an electronic circuit which differentiates, with respect to time, information represented by the amplitude modulation of an a-c carrier wave. A circuit of this type is useful in a-c fire-control computing systems where rate information must be obtained from positional input data. In any fire-control computer, the choice of the particular type of differentiator to be used depends upon the form of the input data. If these data are in the form of d-c signals, accurate electronic differentiating circuits are available. If they are in the form of a-c signals a positional servo coupled to an a-c tachometer is usually used[1] (especially for rectangular co-ordinate fire-control systems). The accuracy of servo differentiators, however, is inadequate for many applications. The electronic circuit to be described offers a means of obtaining, for an a-c fire-control computing system, more accurate differentiation than has so far been obtained using servo differentiators. Basically, the circuit described is a modification of a well-known d-c differentiator circuit which utilizes a Miller integrator in a feedback path, achieving differentiation by the implicit function technique.[2] A magnetically keyed demodulator[3,4] has been incorporated into the design so that the d-c circuit can be used in an a-c computing system. This demodulator circuit has excellent linearity characteristics and a wide operating range, making the electronic differentiator design feasible for many applications.

To confirm the validity of the theoretical considerations outlined in this paper, an experimental differentiator unit for a typical fire-control application has been designed and built. This unit employs the afore-mentioned magnetic demodulator, and magnetic circuits in the stabilization circuit of the integrator amplifier. The results of the tests which were conducted on this unit indicate that an accuracy of ±0.1% can be obtained with satisfactory over-all dynamic performance. To the author's knowledge, the best accuracy obtained using servo differentiators for this application has been about ±1%. The ±0.1% figure therefore represents an improvement in accuracy by a factor of 10.

Description

Fig. 1 is a block diagram of the differentiator circuit showing the interrelation of its four principal components; a d-c integrator, a sampling-type demodulator, a d-c amplifier, and a positional servo. The demodulator and the servo provide the necessary coupling between the a-c system and the d-c differentiator circuit. The demodulator serves two purposes. It converts the input signal from alternating to direct current, and performs the necessary summing of the a-c input signal and the d-c integrator output. The positional servo is of standard design and is used solely to modulate the d-c output of the electronic portion of the differentiator circuit.

In using a servo as a modulator rather than a differentiator, some of the difficulties engendered by the nonlinear characteristics of the feedback potentiometer are not encountered. (In the servo differentiator design, the positional error is differentiated in the same manner as the input signal, which means that the important characteristic of the feedback potentiometer is not its deviation from linearity but rather the slope of linearity error. The fluctuations in the output caused by potentiometer nonlinearity can be made arbitrarily small by reducing the bandwidth of the servo. In reducing the bandwidth, however, the dynamic performance of the system suffers. In many applications, no satisfactory compromise can be found between these two conflicting requirements.) Detailed information on the design of positional servos may be obtained from references 5 and 6.

A block diagram of the d-c portion of the differentiator circuit is shown in Fig. 2. Using the implicit function technique, differentiation is obtained by placing an integrator in the feedback path. If the integrator is assumed to be perfect, the transfer function for the simplified circuit is

$$W(s)=\frac{E_o}{E_i}=\frac{s}{\frac{s}{K_{a1}}+1} \tag{1}$$

The circuit therefore, acts as a differentiator having a bandwidth which depends directly on the gain, K_{a1}, of the d-c amplifier. In a practical system, the determination of the bandwidth is usually based on a compromise between two conflicting requirements, the need for maximum filtering of input noise components, and the need for a rapid settling time for large step inputs. In the experimental differentiator unit, this problem is resolved by using two modes of operation. Under normal operation, when the error signals are small, the bandwidth is approximately a half-cycle per second. When large step input signals occur, however, silicon diodes in the circuit act to increase the bandwidth so that rapid settling time is obtained. When the magnitude of the error voltage is reduced to within its normal limits, the bandwidth reverts to its normal value.

The transfer function of the differentiator circuit based on the more exact block diagram representation, which is shown in Fig. 1, is

$$W(s)=\frac{K_dK_{a1}e^{-\frac{\pi}{\omega_c}s}[RC(K_{a2}+1)s+1]}{RC(K_{a2}+1)s+1+K_dK_{a1}K_{a2}e^{-\frac{\pi}{\omega_c}s}} \tag{2}$$

where ω_c is defined as the computing system carrier frequency. In a practical differentiating circuit the parameters are so chosen that

$K_{a1}>>1$
$K_{a2}>>1$
$K_d\cong 1$
$\omega_c>>$maximum signal frequency

Introducing these relations, the transfer function can be approximated as

$$W(s)\cong\frac{RCs+\left(\frac{1}{K_{a2}}\right)}{\frac{RC}{K_{a1}}s+1} \tag{3}$$

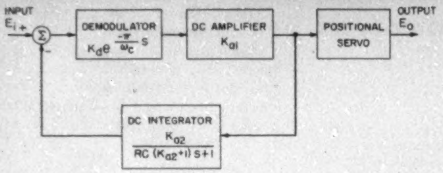
If the gain, K_{a2}, were infinite, equation 3 would be identical in form with equation 1, and the device would act as a perfect differentiator with a finite bandwidth. Because the gain must be finite in a prac-

---

Paper 57-193, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 5, 1956; made available for printing June 17, 1957.

WHEELER X. JOHNSON is with the Radio Corporation of America, Waltham, Mass.

This report is based on a Master of Science Thesis submitted by the author to the Massachusetts Institute of Technology.[14] The author wishes to acknowledge the contributions, made during the course of the thesis research, by Paul R. Johannessen, Dr. John C. Simons, Jr., and Dr. George C. Newton of the Servomechanisms Laboratory. The research reported in this document was supported jointly by the Departments of Army, Navy, and Air Force under Air Force Contract No. *AF 19(122)-458* with Lincoln Laboratory, MIT, and executed under subcontract by the MIT Servomechanisms Laboratory. It is published for technical information only and does not represent recommendations or conclusions of the sponsoring agencies.



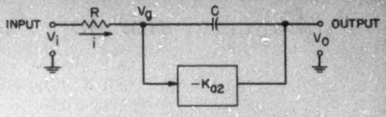


Fig. 3. Miller integrator

Fig. 1 (left). Differentiator block diagram

tical system the differentiation can never be perfect, and the dynamic term in the numerator will introduce an error which will increase with time. This error derives from the imperfect performance of the integrator, and will be discussed at more length in a later section.

In general, the performance, of the differentiator circuit depends upon three factors: the integrator accuracy, the amount of drift, and the amount of internal noise. An estimate of the influence of the first two of these factors can be made by analysis from the theoretical considerations discussed in the section on "Error Considerations." The third must be determined by actual measurement. Of the four principal components of the differentiator circuit, the integrator and the demodulator are of particular interest and are therefore discussed in some detail in the following sections. The remaining two, the d-c amplifier and the positional servo, are of standard design and will not be discussed further.

Integrator Circuit

D-c integrators of the Miller type have been developed to a high degree of accuracy for computing purposes.[7-10] A circuit of this type is used in the experimental electronic differentiator. A block diagram of this circuit is shown in Fig. 3. For the purposes of this report, the circuit transfer function can be assumed to be

$$W_I(s)=\frac{V_o(s)}{V_i(s)}=\frac{K_{a2}}{RC(K_{a2}+1)s+1} \qquad (4)$$

To minimize drift, the gain, K_{a2}, is usually obtained by using a "chopper-stabilized" d-c amplifier.[11] If there are only low frequency components present in the integrator input signal, as is the case in most fire-control problems, the gain can be obtained either by using an a-c amplifier alone or by cascading a-c and d-c amplifiers. This latter arrangement, used in the experimental differentiator unit, is shown in Fig. 4. Because the modulation-demodulation process introduces a time delay into the integrator loop, the use of suitable compensating networks is necessary to obtain stable operation. A stage of d-c amplification is included in order to make full use of the noise-rejection properties of the demodulator and of the compensating networks.

The two silicon diodes and the 22,000-ohm resistor, which are placed in parallel with the 1-megohm integrating resistor, as shown in Fig. 4, reduce the settling time of the differentiator circuit following large step input signals. When a step occurs, the error signal is large; hence, the integrator input signal is also large. As long as the magnitude of the integrator input signal exceeds the "Zener" voltage of the silicon diodes, the diode impedance is very low and the integrating resistance is 22,000 ohms instead of 1 megohm. The integrator time constant, and hence the settling time of the differentiator circuit, is correspondingly reduced. When the magnitude of the integrator input signal drops below the the magnitude of the diode Zener voltage, the diode impedance becomes very high (on the order of 100 megohms) and the integrating resistance again becomes 1 megohm.

The Magnetic Circuits

The experimental differentiator unit utilizes a magnetic modulator and two magnetic demodulators. One demodulator is located in the main loop. The modulator and the second demodulator are employed at the input and the output of the a-c amplifier used in the integrator; see Figs. 1 and 4. The basic mode of operation of these two types of magnetic circuit is similar. Each circuit embodies a gating section, which acts as a switch between the input and output terminals and a driving section, which controls the operation of the switch.

DEMODULATOR

The operation of the demodulator can best be analyzed by considering the basic circuit shown in Fig. 5. The driving section of the circuit causes the switch to close for a short period of time during the positive peaks of the a-c reference voltage wave form. While the switch is closed, the output capacitor, C_o, charges to the peak value of the input voltage. At all other times, the switch is open and the capacitor voltage remains constant. The output, therefore, is a direct voltage having an amplitude equal to the peak amplitude of the a-c input voltage.

The demodulator circuit is shown in Fig. 6. The gating section acts as the "switch" which is closed when the two silicon diodes are conducting current. Conduction is controlled by a series of

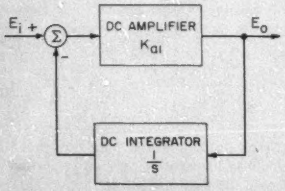


Fig. 2. Simplified differentiator block diagram

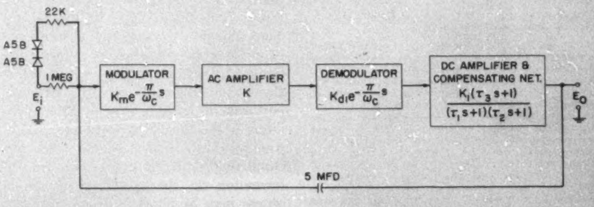

22K
A5B
A5B
1 MEG
E_i
MODULATOR
$K_m e^{-\frac{\pi}{\omega_c}s}$
AC AMPLIFIER
K
DEMODULATOR
$K_{d1} e^{-\frac{\pi}{\omega_c}s}$
DC AMPLIFIER & COMPENSATING NET.
$\frac{K_1(\tau_3 s+1)}{(\tau_1 s+1)(\tau_2 s+1)}$
E_o
5 MFD

Fig. 4. Experimental integrator

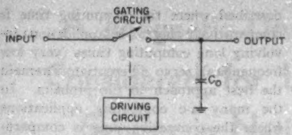


Fig. 5. Basic keyed demodulator circuit

pulses, produced by the driving section[12] and appearing on the secondary windings of reactor II. When a pulse appears the diodes conduct current, and the bias capacitors C_1 and C_2 are charged with the polarity shown in Fig. 6. During the interval between pulses, the charge on the bias capacitors prevents the diodes from conducting. The pulses are timed to occur at the positive peaks of the a-c input voltage waveform so that the desired switching action is obtained.

The demodulator cannot distinguish between a-c and d-c input voltages because it is a sampling device that investigates the magnitude of the input signal at only one point in the a-c voltage wave form. If an a-c signal and a d-c signal are superimposed at the demodulator input, the output is a d-c signal equal to the sum (or the difference) of the two input signals. This summing ability of the demodulator is utilized in the differentiator circuit to obtain the difference between the a-c input signal and the d-c output of the integrator circuit.

The sampling action has the disadvantage that it makes the demodulator very sensitive to high frequency noise components which may appear on the crests of the input signal wave form. This is not a serious difficulty in the differentiator circuit, however, because the objectionable high-frequency noise components can be easily eliminated by suitable filtering of the signal before it reaches the demodulator.

MODULATOR

The basic modulator circuit is shown in Fig. 7. Here, the switch arm changes position every half-cycle of the a-c reference voltage wave, remaining in one position for a full half-cycle. The output is thereby connected on alternate half-cycles to the input and to ground, producing a square wave which has a peak-to-peak amplitude equal to the amplitude of the d-c input voltage, and the same frequency and phase as the a-c reference voltage.

A schematic diagram of the modulator circuit is shown in Fig. 8. The transistor circuit in the driving section[13] produces a square wave on the secondary windings of transformer T_1. When the polarity is as indicated in Fig. 8, diodes D_2 and D_3 conduct current, and diodes D_1 and D_4 are held cut off. Under these conditions, and output terminal is effectively grounded. When the square wave reverses polarity, the pattern of conduction in the diodes is also reversed, and the output terminal is effectively connected to the input terminal. In operation the square wave reverses polarity every half-cycle of the reference voltage, producing the desired switching action.

Error Considerations

INTEGRATOR ACCURACY

The magnitude of the error in the ouput of the differentiator depends primarily on the accuracy of the Miller integrator. The differentiator output (which serves as the input to the integrator) approximates a step when a ramp input signal is applied. The fractional error (Fractional Error=|Correct Value−Actual Value|/Correct Value) in the integrator output for a step input can be determined from the following expression:[10]

$$\text{Fractional Error} = \frac{1}{2}\,\frac{t}{RC\,K_{a2}} \qquad (5)$$

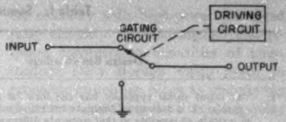


Fig. 7. Basic keyed modulator circuit

The error is smallest when the gain, K_{a2}, is high and the computing time, t, is short.

DRIFT

In using the electronic differentiator circuit as part of an a-c computing system, it is normally desirable to calibrate the circuit so that for a zero rate input, the output voltage is also zero. During subsequent operation, variations in the characteristics of the circuit components may cause the calibration to change so that the output voltage is no longer zero for a zero rate input. Instead, the output voltage has some finite value which represents an error in differentiation. This phenomenon is commonly known as "drift" and must be carefully controlled if high accuracy is to be obtained using the electronic differentiator circuit.

To analyze the effect of component characteristic variations in various parts of the circuit, drift signals can be inserted into the block diagram representation at suitable points, and the variations in the output caused by these signals can be calculated. It can be shown that the portion of the output signal caused by drift varies in proportion to the rate of change of drift voltages in the main loop amplifier, and to the magnitude of drift voltages in the integrator amplifier. In most applications, the drift rate in the main loop amplifier is not great enough to have an appreciable effect on the output. The error from this source, therefore, is negligible.

Great care must be taken, however, to

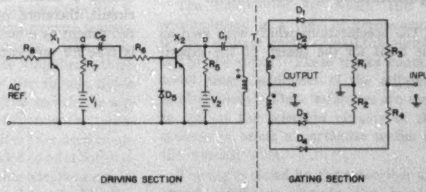


Fig. 8. Schematic diagram of magnetic modulator

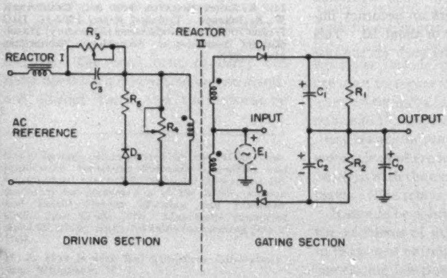


Fig. 6 (left). Schematic diagram of magnetic demodulator

Table I. Summary of Test Results

Design Specifications	Results of Tests on Experimental Unit*
I. The input signal typically has the form of a step followed by a ramp. It is desired to dissipate the transient caused by the step as quickly as possible so that accurate differentiation of the ramp signal can be obtained	
A. Ramp Signal	
1. The ratio between the maximum and minimum detectable rates of change of signal magnitude must be 500 to 1 or greater	Rates of from ±0.004 volt to ±2 volts per second are detectable
2. The accuracy of the differentiation must be within ±0.1% of the maximum rate	The maximum error in differentiation is less than ±0.002 volt which is equivalent to ±0.1% of the maximum rate
3. The maximum detectable signal which must be handled is best defined in terms of the maximum rate of change of signal magnitude. The maximum signal capability must be such that the maximum rate of change can be maintained without saturation for a period of 30 seconds	The unit will handle signals in the range between +30 volts and −30 volts. The total range, therefore, is 60 volts (2 volts per second × 30 seconds = 60 volts)
4. For ramp inputs which do not saturate the system, the maximum period during which differentiation is required is 2 minutes	The unit maintains the specified accuracy at least 2 minutes
B. Step Signal	
1. The maximum step signal which must be handled has the same magnitude as the signal described under A-3	The maximum step which can be handled is from ±30 volts to ∓30 volts
2. The settling time must be less than 4 seconds. Settling time is defined as the length of time following a step during which the accuracy of differentiation is outside the specified limit	For a maximum step input, the settling time is slightly less than 4 seconds
II. The differentiator must operate in an a-c computing system using a 400-cycle-per-second carrier frequency	

* The voltages given are peak rather than rms values.

insure that the integrator drift voltages are reduced to a level commensurate with the accuracy requirements. As mentioned in the section on the integrator circuit, the chopper-stabilization technique is usually used to control the drift characteristics of the integrator amplifier. The effectiveness of this technique depends upon the fact that the mechanical chopper, which is ordinarily used as the modulator, does not itself produce any drift signals. The magnetic modulator circuit described in this paper does produce drift signals. It has been found, however, that by using reasonable care in the design and construction of the circuit, these signals can be made small enough so that the modulator can be used instead of a chopper in many applications involving the electronic differentiator circuit.

Specifications and Conclusions

The specifications which were used as the basis for the design of the experimental differentiator unit are listed in Table I. These specifications themselves are based on conditions representative of those actually encountered in fire-control computer applications. The specified input signals admittedly are idealizations of the actual field signals. Their use, nonetheless, provides a good indication of the differentiator performance, and they are easily simulated in the laboratory.

A summary of the results of the tests which were made on the unit, confirming its conformance with the specified requirements, is also shown in Table I. In meeting the prescribed specifications, the electronic unit has exhibited a marked improvement in performance over that which has so far been obtained using servo differentiators. For a servo which meets the dynamic performance requirements, the best accuracy which the author knows to have been so far obtained is about ±1%. The electronic circuit, therefore, offers an accuracy improvement by a factor of about 10. This is true for applications such as the one described where the computing time is comparatively short. In applications involving long computing times (very low frequencies), servo differentiators remain the best approach to the problem. In the many a-c computing applications where the computing time is comparatively short, the increased accuracy obtainable with the electronic differentiator circuit should recommend its use in spite of the added complexity.

References

1. A DUAL-BANDWIDTH A-C SERVO DIFFERENTIATOR, **M. L. Shechet.** *M.S. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., 1954.

2. ELECTRONIC INSTRUMENTS (book), **I. A. Greenwood, J. V. Holdam, Jr., D. MacRae, Jr.** McGraw-Hill Book Company, Inc., New York, N. Y., MIT Radiation Laboratory Series, vol. 21, 1948, pp. 15–21.

3. MAGNETICALLY KEYED, PHASE-SENSITIVE DEMODULATORS, **Robert B. Mark, Wheeler X. Johnson, Paul R. Johannessen.** *AIEE Transactions*, vol. 76, pt. I., Mar. 1957, pp. 1–6.

4. HIGH PERFORMANCE DEMODULATOR FOR SERVOMECHANISMS, **K. R. Schreiner.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. II, Oct. 3–5, 1946, pp. 393–403.

5. PRINCIPLES OF SERVOMECHANISMS (book), **G. S. Brown, D. P. Campbell.** John Wiley & Sons, Inc., New York, N. Y., 1948.

6. AUTOMATIC FEEDBACK CONTROL SYSTEM SYNTHESIS (book), **J. G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

7. WAVEFORMS (book), **B. Chance, V. Hughes, E. F. MacNichol, Jr., D. Sayre, F. C. Williams.** McGraw-Hill Book Company, Inc., New York, N. Y., MIT Radiation Laboratory Series, vol. 19, 1949.

8. ELECTRONIC ANALOG COMPUTERS (book), **G. A. Korn, T. M. Korn.** McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

9. LONG TIME ANALOG MEMORY, **J. T. Torn, S. Katz, R. N. Close.** *Proceedings*, National Electronics Conference, vol. 10, 1954.

10. A LOW-DRIFT INTEGRATING AMPLIFIER, **W. R. Porter.** *Technical Memorandum 6897-TM-18*, DIC Project 6897, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1954.

11. STABILIZATION OF WIDE-BAND DIRECT-CURRENT AMPLIFIERS FOR ZERO AND GAIN, **E. A. Goldberg.** *RCA Review*, Princeton, N. J., vol. 11, 1950, p. 296.

12. PULSE RELAXATION AMPLIFIER—A LOW LEVEL AMPLIFIER, **R. E. Morgan, J. B. McFerran.** *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 345–50.

13. JUNCTION TRANSISTORS USED AS SWITCHES, **R. L. Bright.** *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

14. A DIFFERENTIATOR FOR A-C COMPUTERS, **W. X. Johnson.** *Technical Report 7002-12*, DIC Project 7002, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

Vol.
Pt. 1
1958

A Magnetic-Drum Storage System Considered for Use as a Common Sender in Nation-Wide Dialing

H. F. MAY
MEMBER AIEE

THE inauguration of direct-distance dialing by the customer requires use of telephone numbers consisting of ten digits. Thus a telephone subscriber in San Francisco will dial 212 CH 3-1000 in order to reach the Bell Telephone Laboratories in New York City. The first three digits are the code for the metropolitan New York area, while the remaining seven digits comprise the telephone number dialed by subscribers within the metropolitan area. To handle calls involving ten digits, the digit-storing capacity of many existing offices will have to be extended and various methods of accomplishing this are under consideration.

The current state of the computer art suggested the investigation of electronic techniques as adjuncts to telephone switching, and a laboratory project was initiated to devise an electronic solution to the problem of extending the digit-storing capacity of no. *1* crossbar and panel sender offices. For economic reasons, a relay solution was eventually chosen for use in the field.

The master drum sender, or MDS, is an auxiliary common (control) register utilizing a magnetic-drum memory to provide 11-digit registering capacity for 156 senders. It provides the circuitry for counting and registering the dial pulses occurring simultaneously in 156 senders and also for outpulsing these digits in multifrequency code after complete numbers have been stored.

Through the use of electronic switching techniques employing transistors and vacuum tubes, the speed is such that a single MDS may be time-shared by 156 senders. The equipment to carry out these operations is mounted on two standard 11-foot racks and the laboratory model of this equipment is shown in Figs. 1 and 2. The motor-driven unit in the background is the magnetic drum. The far bay contains the vacuum-tube read and write amplifiers associated with the magnetic drum and the transistor circuitry which carried out the logical operations required of the MDS. The near bay contains relay equipment used to establish connections with the 156 senders and to apply the multifrequency tones to the selected sender. The actual senders are part of a no. *1* crossbar laboratory system located in an adjacent room. As the model now stands, it contains approximately 500 transistors, 2,000 diodes, 69 electron tubes, and 68 relays.

To convey some idea of the equipment used, the following illustrations are included. Fig. 3 is the magnetic drum used in the laboratory model. It is a drum similar to that used in the DIAD (Drum Information Assembler and Dispatcher) system, another electronic switching system[1] developed by the Bell Telephone Laboratories. The magnetic drum intended for field use is shown in Fig. 4, and this unit is designed for rack mounting. The vacuum-tube circuitry which consists of read and write amplifiers and a clock-pulse generator is shown in Fig. 5. The clock-pulse rate is 60 kc and the clock-pulse width is 2.5 μsec (microseconds). The transistor and diode circuitry are mounted on plug-in printed circuit cards and a typical section of this circuitry is shown in Fig. 6. Detailed views of typical cards are shown in Fig. 7. Both junction- and point-contact transistor are used in this circuitry. The package types include pulse amplifiers which are used to regenerate the signals, diode AND and OR gates used to switch the signals and flip-flops used for memory. Auxiliary transistor amplifiers are used to operate relays where needed to establish connections with the existing telephone equipment. The relays are shown in Fig. 8.

It should be pointed out here that one of the problems of utilizing electronic logic in telephone switching is the difficulty of associating low-power high-speed transistor circuitry with high-power low-speed relay circuitry. To accomplish this requires considerable quantities of intermediate relay storage which poses an economic stumbling block to introducing electronic switching into existing relay offices.

Since some of the equipment utilized in MDS has been discussed, the organization of the system is now explained by reference to Fig. 9, which is a block diagram of MDS. In this diagram, the originating subscriber in a no. 1 crossbar office is shown connected to one of three senders by means of the usual X-bar switching equipment. It is through the originating sender that the MDS picks up dial pulses and sends out multifrequency tones, representing the called number, to the terminating office, which in this case is designated as a no. 5 crossbar office. In the sender there is a relay which follows dial pulses, and the contact of this relay is connected into an electronic scanner. The scanner examines the contact of the pulsing relay in each of the 156 senders every 10 milliseconds. During this short interval, a dial pulse cannot be completed so that the scanner will detect every operation of the pulsing relay. The results of this scanning operation are interpreted by the sender space control as off-normal, preliminary pulse, dial pulse, end of pulse train, etc., and the interpretation is stored on the magnetic drum in a space assigned to each sender. This space is physically 1 inch by 1/4 inch, Fig. 10, and contains 84-unit magnetizable areas known as cells. Within each cell a single bit or pulse can be stored. The drum circuit controls the reading and writing operations on the drum.

The operation of the scanner is synchronized with the rotation of the drum by means of the address circuit so that when a sender pulsing relay is being scanned, the corresponding sender space on the magnetic drum is passing under the reading and writing heads.

The mechanism by which this is accomplished is a pair of permanently written address tracks around the periphery of the drum. These tracks contain serial binary addresses of the adjacent sender spaces. These addresses are used to set up the selection circuits of the scanner to select a particular sender in the crossbar office.

The digits dialed by the subscriber are stored simultaneously in the regular crossbar sender and in the corresponding sender space on the drum.

When sufficient information has been stored in any sender space, e.g., seven

---

Paper 56-666, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted April 20, 1956; made available for printing July 9, 1957.

H. F. MAY is with Bell Telephone Laboratories Inc., Whippany, N. J.

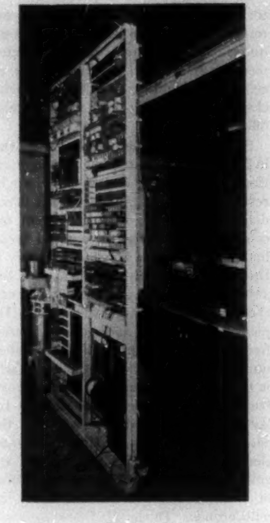


Fig. 1. MDS model, front view

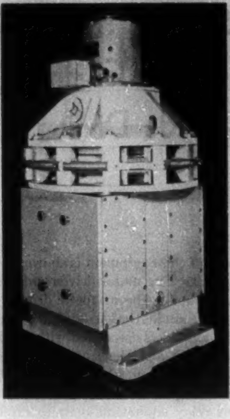


Fig. 3. DIAD magnetic-drum store

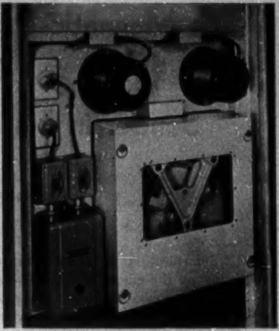


Fig. 4. MDS magnetic-drum store

digits, to require action by MDS, the transmitter control swings into action. The transmitter control is continuously monitoring the sender spaces, and it is the function of the transmitter control to take information out of the sender spaces and to control multifrequency outpulsing to the crossbar senders.

All actions are performed by MDS through the senders in the crossbar office. Consequently, the signal level and time scale of the MDS control are dictated by the relay requirements of the crossbar office. The decision-making and memory elements of MDS are electronic in nature and function at speeds measured in microseconds, while the relay operations take place in milliseconds. Because of this thousandfold ratio in speed, the transmitter control carries out its decisions through intermediary units known as transmitters. These transmitters contain both relays and electronic equipment. They receive information at electronic speeds and execute operations at relay speeds through the senders of the crossbar office. Conversely, the transmitters receive information from the senders at relay speeds and then transmit this information at electronic speeds to the transmitter control. Essentially the transmitters are intermediate storage devices for conversion between electronic and relay speeds and signal levels.

The number of transmitters provided in an MDS unit would be determined by the traffic in the central office. It has been

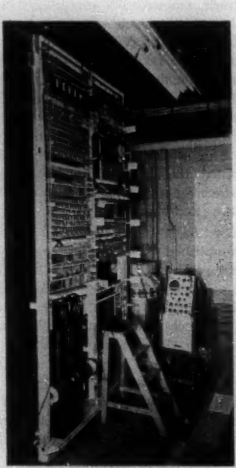


Fig. 2. MDS model, rear view

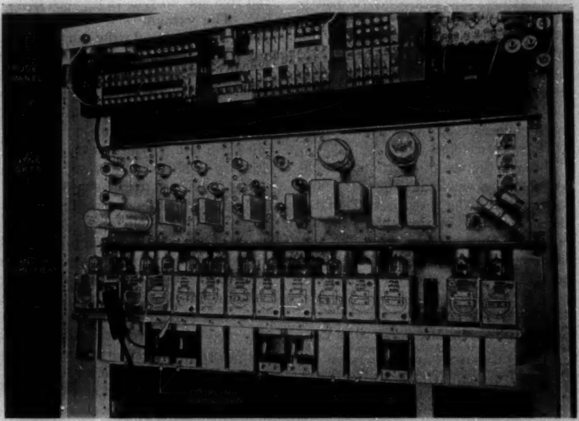


Fig. 5. Drum amplifiers and clock generator

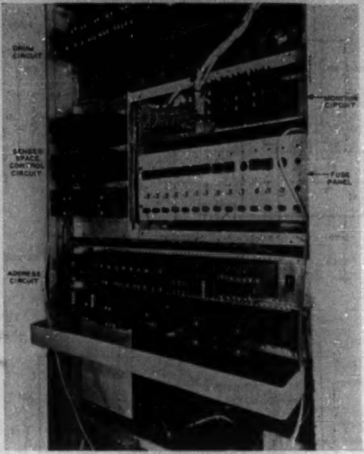


Fig. 6 (left). Transistor and diode plug-in circuitry

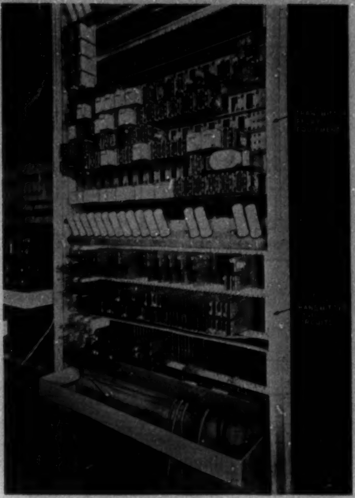


Fig. 8 (right). Transmitter bay

estimated that six transmitters could serve a group of 156 senders in a typical office. In the model, a single transmitter suffices for testing. Where a group of transmitters is controlled by the transmitter control, the first idle transmitter receives the order from the transmitter control. Successive orders for the same sender may be executed by any idle transmitter.

It would now seem appropriate to describe in detail several of the components and their circuits. The magnetic drum is a nickel-cobalt-plated steel cylinder, rotating at 1,500 rpm. Tracks on the surface of the drum are spaced 1/16 inch apart, and the lamination pileup of the head is 0.050 inch wide. Packing factor is about 60 per inch. Pulses are recorded on a saturated background. A unique feature of this magnetic recording system is that single-pass reading and writing is employed. With single-pass operation a cell can be read, the information therein combined in gating circuits with information from other tracks or outside sources, and new information written in the same cell, all while the cell is passing under the head. The reason why this can be accomplished is indicated in Fig. 11. The voltage read back from a recorded pulse is considerably wider than the current pulse which did the recording. This may be caused by leakage flux around the head gap during recording, or by the spread of flux beyond the magnetized area due to the permeability of the magnetic surface. The result is that it is possible to obtain a useful signal from a reading amplifier 2.5 μsec before the write pulse occurs. As a consequence of single-pass operation, the information in every cell in a track can be processed in a single revolution of the drum, and this results in high traffic-handling capacity with a minimum of equipment.

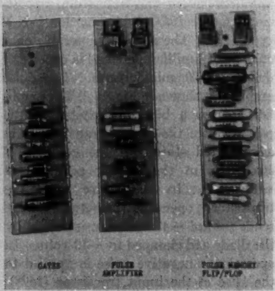


Fig. 7. Typical plug-in cards

To achieve single-pass operation, the reading amplifiers must have a wide passband and a short recovery time after overload. The reading amplifier shown in a previous picture is a plug-in package containing three no. *396A* electron tubes. The circuit is shown in Fig. 12. The first two tubes comprise a linear amplifier with a stabilized gain of 750. The third tube is a threshold stage biased beyond cut-off and a phase inverter. The linear amplifier raises the head voltage (50 millivolts P-P) to 40 volts which is applied to the threshold stage. This stage rejects background noise but responds to the positive-going half-cycles from the linear amplifier. The phase inverter provides a complementary output so that signals are available for both "zero" and "one."

The writing amplifier is similar in external appearance to the reading amplifier and also utilizes three *396A* electron tubes. This circuit is shown in Fig. 13. Two of the tubes operate as blocking oscillators, one serving to generate a "write 1" pulse and the other generating the "write 0" or erase pulse. The oscillators produce a 0.25-ampere current pulse in the head when triggered from the buffer amplifier which is the input stage.

The reading and writing amplifiers are coupled to the magnetic head through a transient suppressing network. This network damps the trailing edge of the writing pulse so that any tendency to ring is damped out before the next cell is reached. By keeping down the amount of reactance in the coupling circuits in the read amplifier, its recovery time is short enough so that it is ready to read within 7 μsec after the write pulse.

The major portion of the MDS circuitry consists of germanium diodes

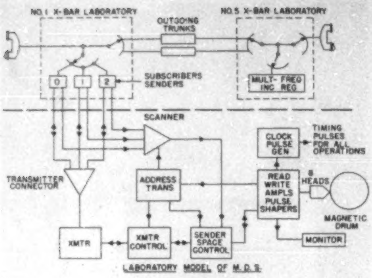

NO. 1 X-BAR LABORATORY
NO. 5 X-BAR LABORATORY
OUTGOING TRUNKS
SUBSCRIBERS SENDERS
0
1
2
MULT-FREQ INC REG
SCANNER
TIMING PULSES FOR ALL OPERATIONS
CLOCK PULSE GEN
TRANSMITTER CONNECTOR
ADDRESS TRANS
READ WRITE AMPLS PULSE SHAPERS
8 HEADS
MAGNETIC DRUM
XMTR
XMTR CONTROL
SENDER SPACE CONTROL
MONITOR
LABORATORY MODEL OF M.D.S.

Fig. 9. Block diagram of MDS

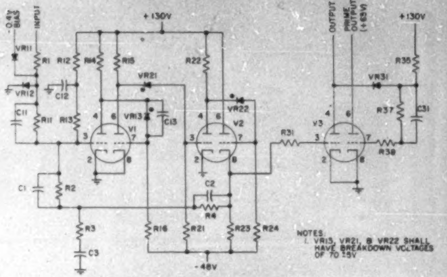

-0.4V BIAS
INPUT
+130V
OUTPUT
PRIME OUTPUT (+63V)
+130V
VR11
R1
R12
R14
R15
R22
R35
VR12
C12
VR21
VR31
C11
R11
R13
VR13
C13
VR22
R37
C31
V1
V2
R31
V3
R38
C1
R2
C2
R4
R3
R16
R21
R23
R24
C3
-48V
NOTES
1. VR13, VR21, & VR22 SHALL HAVE BREAKDOWN VOLTAGES OF 70 ±5V

Fig. 12. Read amplifier schematic

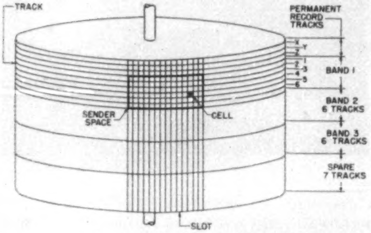

TRACK
PERMANENT RECORD TRACKS
X
Y
BAND 1
BAND 2 6 TRACKS
BAND 3 6 TRACKS
SPARE 7 TRACKS
SENDER SPACE
CELL
SLOT

Fig. 10. Drum layout showing sender space

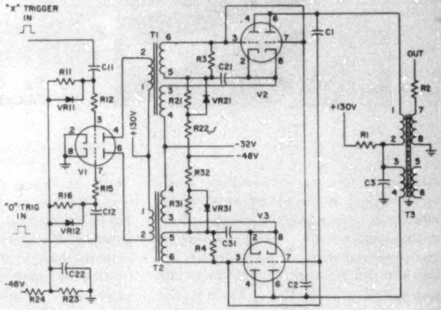

"X" TRIGGER IN
T1
C1
OUT
R11
C11
R3
C21
VR11
R12
R21
VR21
V2
R2
+130V
R22
+130V
R1
-32V
-48V
T3
V1
R32
C3
R15
"O" TRIG IN
R16
C12
R31
VR31
VR12
V3
C31
R4
T2
C22
C2
-48V
R24
R23

Fig. 13. Write amplifier schematic

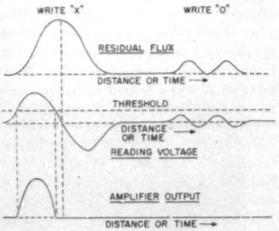


Fig. 11. Waveforms in reading from drum

and transistors which together with associated resistors are mounted on plugin-cards such as shown in Fig. 7. These cards are produced by conventional etched circuit techniques. Diode AND and OR gates are used to carry out the logical operations in a manner usually employed in computers. Yokelson and Ulrich[2] describe the details of design of these gates.[3] However, to complete the description, the several types of gates used in MDS are shown in Fig. 14. AND gates consist of a load resistor fed from negative battery, and one diode for each input. These inputs are at potentials of either −15 (zero) or −30 (one), and the most positive input determines the output. All inputs must be at −30 (one) for the output to present a "one" signal.

OR gates consist of a load resistor fed from ground and a diode for each input. Potentials applied to the inputs are the same as for AND gates but the reversed polarity of the diodes now allows the most negative input to determine the output. Thus a single −30-volt (one) potential produces a "one" output.

A rather special form of gate which can provide high-current output pulses at a low duty factor is the transmission-type AND gate. This gate is used to couple a low-power flip-flop to pulse amplifiers. In this unit, the condenser *C* is discharged to a −30-volt potential by the clock pulse. If a −30-volt potential is maintained between clock pulses by the flip-flop output, the condenser charge remains unchanged and the next clock pulse produces no change in charge. If the flip-flop output rises to −15 volts between clock pulses, the condenser will charge to this potential through high resistance R_s. The next clock pulse will discharge the condenser and produce an output current pulse. The transmission gate integrates a relatively small current over a time interval of about 7 μsec and delivers a pulse of 2-μsec duration and correspondingly higher current.

Pulse amplifiers are used to regenerate the various signals in MDS. These amplifiers utilize two transistors, a point-contact type *1698* and a junction P-N-P type *1778*. The amplifier operates on a current of 1 milliampere and is capable of delivering 80 milliamperes. The pulse amplifier is shown in Fig. 15. The amplifier acts as a bistable switch. In the passive state the output transistor is in a saturated state and the collector is essentially at the emitter potential of −12 volts. In the active state the transistor is open and the output voltage tends toward −48 volts but is caught by the diode and clamped to −30 volts. In operation a negative pulse is applied to the base of the input transistor (*1698*). Conduction from emitter to base turns the transistor on and because of the

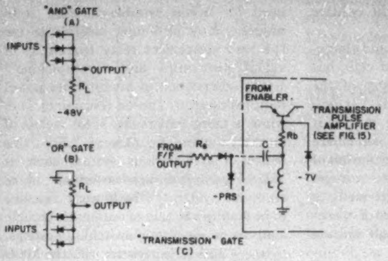

"AND" GATE
(A)
INPUTS
OUTPUT
R_L
-48V
"OR" GATE
(B)
R_L
OUTPUT
INPUTS
"TRANSMISSION" GATE
(C)
FROM ENABLER.
TRANSMISSION PULSE AMPLIFIER (SEE FIG. 15)
Rb
FROM F/F OUTPUT
R_s
C
-7V
L
-PRS

Fig. 14 (left). Diode gates

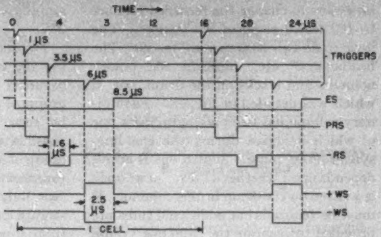

TIME
0
4
8
12
16
20
24 µS
1 µS
3.5 µS
6 µS
8.5 µS
TRIGGERS
ES
PRS
-RS
1.6 µS
2.5 µS
+WS
-WS
1 CELL

Fig. 17 (right). Clock-pulse timing

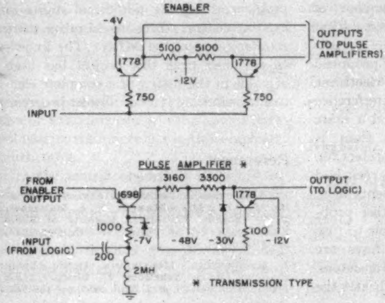

ENABLER
-4V
5100
5100
1778
1778
-12V
750
750
INPUT
OUTPUTS (TO PULSE AMPLIFIERS)
PULSE AMPLIFIER *
FROM ENABLER OUTPUT
1698
3160
3300
1778
OUTPUT (TO LOGIC)
1000
-7V
INPUT (FROM LOGIC)
200
2MH
-48V
-30V
100
-12V
* TRANSMISSION TYPE

Fig. 15 (left). Enabler and pulse amplifier

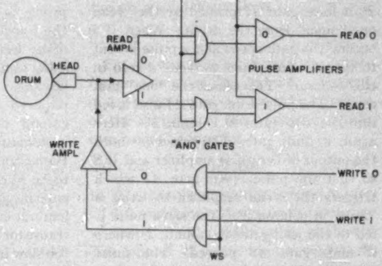

RS
ES
READ AMPL
0
1
HEAD
DRUM
PULSE AMPLIFIERS
READ 0
READ 1
WRITE AMPL
"AND" GATES
WRITE 0
WRITE 1
WS

Fig. 18 (right). Logic diagram of read-write circuitry

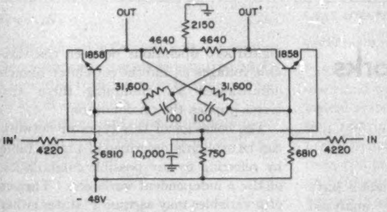

OUT
OUT'
2150
4640
4640
1858
1858
31,600
31,600
100
100
IN'
4220
6810
10,000
750
6810
4220
IN
-48V

Fig. 16 (left). Flip-flop

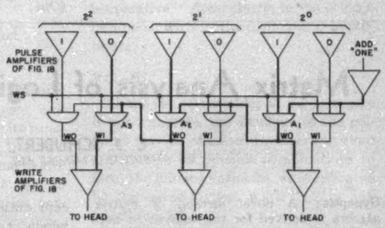

2^2
2^1
2^0
PULSE AMPLIFIERS OF FIG. 18
1
0
ADD "ONE"
WS
A_3
A_2
A_1
WO
WI
WRITE AMPLIFIERS OF FIG. 18
TO HEAD
TO HEAD
TO HEAD

Fig. 19 (right). Logic diagram of adder

trigger characteristics of the point-contact transistor, it remains in a conducting state until turned off. When the trigger stage is on, the output stage is turned off. The trigger stage remains on until the emitter potential is reduced below the base potential of −7 volts. The emitter is supplied from a clock-pulse amplifier called an "enabler." For the duration of the clock pulse, the emitter of the point-contact transistor is connected to −4 volt. At the end of the clock pulse the emitter is supplied from −12 volts through 5,100 ohms. This drops the emitter below the base potential and turns off the transistor. When the point-contact transistor turns off, the output transistor turns on and output voltage rises to −12 volts.

Another basic circuit used in MDS is the flip-flop which is used as a memory element. These are bistable symmetrical circuits utilizing two N-P-N transistors, type *1858*. The circuit of a flip-flop is shown in Fig. 16. The flip-flop operates on a 5-volt negative pulse and its output signal levels are approximately −17 and −36 volts. Since the flip-flop can supply very little current it is always used with a transmission-type AND gate as previously described.

As a simple example of the type of circuitry in which these component packages are used, consider the logical operations employed in adding a pulse to a binary number stored in the drum. This will explain the advantages of single-pass reading and writing. It is first necessary to understand the clock-pulse timing and this is shown in Fig. 17. The important thing to note is the existence of three separate clock pulses during each cell space or cycle. The read-synchronizer or *RS* pulse precedes the write-synchronizer *WS*, and the enable pulse *ES* includes both *RS* and *WS* within its envelope. All signals read from the drum are gated by *RS*. All writing is gated by *WS*. The signals read from the drum and gated by *RS* are used to operate pulse amplifiers which lock and remain active until the enable pulse ceases. These basic operations are shown in Fig. 18.

Assume now that on the drum in a particular row of three cells the binary number *011* is stored and it is desired to add one pulse to this number. It is necessary to change the recorded number to *100* which requires two carries. The logic to do this is shown in Fig. 19. The headings 2^0 etc., indicate the binary columns and tracks on the drum, one of which was detailed in Fig. 18. Each binary column has two pulse amplifiers one of which is active during the enabling-synchronizer pulse. Which one is active depends upon whether a "zero" or a "one" is written on the drum in this column. It must be realized that additional gates are provided to generate the "add one" pulse at the instant that the binary number *011* is being read. When the "add one" pulse arrives at the pair of AND gates in column 2^0, it finds gate *A1* primed by the "read one" signal from the drum. When *WS* occurs, the gate opens and a pulse is sent to the write amplifier to write a zero in this column. This erases the "one" first read. The pulse from gate *A1* also is fed into the AND gates in column 2^1. Here again it finds gate, *A2* primed by both the output of the pulse amplifier and *WS* so that the pulse opens gate *A2* which triggers the write amplifier to write a "zero" in column 2^1. The same pulse is fed to the AND gates in column 2^2 where it finds gate *A3* primed. The pulse opens gate *A3* which triggers the write amplifier to write a "one" in column 2^2.

The entire action of reading and changing the content of a group of cells has occurred during a single passage of the group of cells under the heads. It will be seen that this technique makes it possible for MDS to service an entire group of senders during each revolution of the drum without intermediate storage. Not only is the need for intermediate storage eliminated, but the speed of operation is increased so that a single unit of MDS can serve 156 senders.

The operation of the MDS in the laboratory brought out a number of problems which had to be solved. First the transients developed in relay circuits many feet from the MDS rack produced false signals in the low-level electronic circuits. To reduce this interference required a more liberal use and a more careful design of suppressor than is customarily used for contact protection. Power leads entering the MDS rack had to be filtered to prevent transients from entering on these leads. Another problem not completely solved is how to fuse transistor circuits. Ordinary fuses are too slow in blowing to protect transistors. Experience in the laboratory indicates the need for better trouble-shooting techniques. It is no longer possible to use the very convenient relay techniques of visual observation and "toothpicking." Every observation of microsecond pulses requires carefully chosen test points, and, what is more important, some means of knowing when to observe. For this reason a synchronous system such as MDS has advantages over an asynchronous system. Much work remains to be done in the field of automatic trouble analysis on electronic switching systems.

In conclusion, experience with the MDS indicated that a relay solution to the problem of storing additional digits in existing central offices would prove more economical than the MDS. The knowledge gained from this project has been of value in the design of a complete electronic switching system now in progress.

References

1. AN AUTOMATIC TELEPHONE SYSTEM EMPLOYING MAGNETIC DRUM MEMORY, **W. A. Malthaner, H. E. Vaughn.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Oct. 1953, pp. 1341–47.

2. ENGINEERING MULTISTAGE DIODE LOGIC CIRCUITS, **B. J. Yokelson, W. Ulrich.** *AIEE Transactions*, vol. 74, pt. 1, Sept. 1955, pp. 466–75.

Matrix Analysis of Logical Networks

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: A novel method of matrix algebra is derived for the analysis of networks representing logical systems too complex for conventional approaches using Boolean algebra. It is shown that truth tables of system blocks may be treated as matrices and that logical operations on such matrices can be performed efficiently.

THE ANALYSIS of logical networks employing the techniques of Boolean algebra becomes impracticable when the logical function is dependent on many variables. In order to analyze logical networks, several simple computers[1–4] have been designed. With the matrix concept outlined, up to ten variables have herein been mathematically analyzed without the aid of a computer in a time shorter than that required to set up a computer program for the same problem.

Logical Networks

Any network performing logical operations may be represented by a set of voltage sources and sets of circuit elements. Voltage and impedance are understood to be limited to two states only to express ONE and ZERO conditions. The total current I follows the general equation (see Fig. 1)

$$I = U_1Y_1 + U_2Y_2 + U_3Y_3 + U_4Y_4 \qquad (1)$$

wherein the admittances of the branches might be functions of the independent logical variables as follows:

$$Y_1 = Y_1(X_1, X_2, X_3, X_4) \qquad (2)$$

The products in equation 1 involve logical AND operations between the variable voltages U_1 and the correlated branch admittances Y_1. Summing these AND terms implies the OR condition.[5]

The behavior of this types of network can be explicitly described in a truth table by referring to any possible combination of the p independent variables. The set of p variables may assume 2^p states either in an AND or an OR condition, called "minterm" and "maxterm," respectively. A minterm N/maxterm M is any one of the 2^p distinct AND/OR conditions of all p variables in which these variables are or are not complemented in any of the 2^p possible ways.[6] The suffix of N or M is the decimal equivalent of the binary number formed by writing ZERO for each complemented variable and ONE for each uncomplemented variable, in the order of their appearance.

No further discrimination between voltage and impedance as logical variables is necessary; therefore all independent variables are denoted by X_i in opposition to Y and Z, for logical functions or dependent logical variables, as shown in the diagram and related truth table of Fig. 2. The state of variables X_i is defined by the minterms N; sums of certain

---

Paper 57-673, recommended by the AIEE Committees on Basic Sciences and Computing Devices and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 22, 1957; made available for printing September 6, 1957.

E. J. SCHUBERT, previously with the Westinghouse Electric Corporation, Baltimore, Md., is now with Burroughs Corporation, Paoli, Pa.

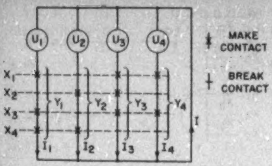


Fig. 1. Contact network performing logical operations

minterms form functions Y. In general any function Y of p variables can be expressed as a polynomial of all minterms with appropriate coefficients F_q.

$$Y=F_0N_0+F_1N_1+\ldots+F_qN_q \quad q=2^p-1 \tag{3}$$

wherein the coefficients F_i have the values ONE or ZERO according to the appropriate column for the function Y in the composite truth table.

In the matrix operations described in the following the conventional matrix notation is supported by index notation as a supplement. Using the Einstein convention, i denotes the dummy index between $q=0$ and $q=2^p-1$, z is the fixed index. Capital letters designate the individual terms of the corresponding matrices in index notation.[7]

$$Y_z=F_{zi}N_i \tag{4A}$$

Multiplication between matrices implies the appropriate transposed matrix, which is indicated by a suffix t only when needed to avoid ambiguity.

In conventional matrix notation all dependent variables of a network are described by

$$\mathbf{y}=\mathbf{fn} \tag{4B}$$

if the matrices are defined as

$$\text{minterm matrix } \mathbf{n}=(N_0N_1N_2\ldots N_q) \tag{5}$$

$$\text{function matrix } \mathbf{f}=(F_0F_1F_2\ldots F_q) \tag{6}$$

OR Operation

The sum of minterms is by definition an OR operation. Because functions Y are expressed as sums of minterms, any sum of functions Y again will be a sum of minterms. In matrix notation the addition yields

$$Y_1+Y_2=(F_i+G_i)N_i \tag{7A}$$

or

$$\mathbf{y}_1+\mathbf{y}_2=\mathbf{fn}+\mathbf{gn}=(\mathbf{f}+\mathbf{g})\mathbf{n} \tag{7B}$$

The block diagram of Fig. 3 shows the two logical networks with the function matrices f_1 and f_2. Their outputs y_1 and y_2, are buffered into an OR gate. The equivalent circuit, shown in part B of Fig. 3, symbolizes the required function obtained by matrix algebra.

AND Operation

In Fig. 4(A) two logical networks feed an AND gate. To obtain an output y implies that both the y_1 and y_2 inputs into the gate are ONE. Because the state of the output depends on the coefficients of the function matrices, f and g, y can be ONE only if the coefficients in both matrices with identical suffix are ONE. All other minterms must be inhibited. The following relations apply:

$$\mathbf{y}_1=\mathbf{fn} \qquad \mathbf{y}_2=\mathbf{gn} \tag{8}$$

$$Y=(F_iG_i)N_i \tag{9A}$$

$$\mathbf{y}=\mathbf{y}_1\mathbf{y}_2=(\mathbf{fg})\mathbf{n} \tag{9B}$$

Negation

This operation implies that each F_{zi} term must be substituted by its complement F'_{zi}. According to elementary definitions

$$F_{zi}+F_{zi}'=1 \qquad \ldots\text{NOT} \tag{10A}$$

$$F_{zi}'=1-F_{zi} \tag{10B}$$

$$Y_z'=(1-F_{zi})N_i \tag{10C}$$

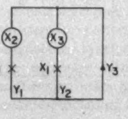
Or in conventional matrix notation

$$\mathbf{y}=(\mathbf{1}-\mathbf{f})\mathbf{n}=\mathbf{f}'\mathbf{n} \tag{10D}$$

As an electrical network is sufficiently described with its impedance tensor, any logical system can be defined by its function matrix. It is understood that the minterm matrix is one-dimensional in any case, comprising only one row or column and identical for any operation performed in a particular system. The function matrix may comprise up to 2^{2p} rows, but in practice only those rows describing utilized outputs are considered.

Function Matrix

In establishing the function matrix, certain rules are pertinent. The columns of the function matrix shall be numbered from 0 at the left up to 2^p-1, corresponding to the dummy index of the minterm matrix. If all variables occur on the inputs of AND gates, the function matrix is directly available as follows.



X_3	X_2	X_1	N	Y_1	Y_2	Y_3
0	0	0	N_0	0	0	0
		1	N_1	0	0	0
	1	0	N_2	0	0	0
	1	1	N_3	1	0	1
1	0	0	N_4	0	0	0
1	0	1	N_5	0	1	1
1	1	0	N_6	0	0	0
1	1	1	N_7	1	1	1

Fig. 2. Logical network and its truth table

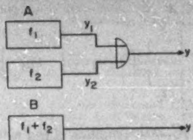


Fig. 3. A—OR condition of two networks. B—Equivalent circuit

RULE I

For the complete set of variables occurring at AND gates, the decimal equivalent of the binary number formed by the state of the variables is the column in which ONE has to be inserted.

$AB'CDE'$ … 10110(b) … 22(d)

In this example an AND gate is set up for all five variables in the system to generate an output for the combination as shown. The number 22 is the decimal equivalent of the binary number corresponding to the state of the inputs for which the gate is operative. Accordingly, in the column numbered with 22 the digit ONE is appropriate.

RULE II

If the set of variables at an AND gate is not complete, the numbers of the columns may be obtained by inserting ZERO and ONE in all possible combinations to form the missing variables, as is indicated by the use of a check mark "√" or any other symbol.

$A\surd C'D\surd$	…	10010(b)	…	18(d)
		10011	…	19
		11010	…	26
		11011	…	27

In columns 18, 19, 26, 27 the digit ONE has to be inserted.

RULE III

A complete set of variables on the input of an OR gate generates a single maxterm, denoted by the suffix similar to minterms. Because the matrix analysis is based on minterms only, a conversion from maxterms into minterms is necessary. As derived by basic Boolean algebra for a particular case:

$$A+B'+C+D+E'=M_{22}=(A'BC'D'E)'=N_9'$$

For the relation of the suffixes of a max-

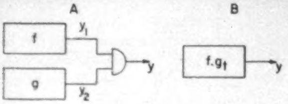


Fig. 4. A—AND condition of two networks. B—Equivalent circuit

term M_i and its equivalent complement of a minterm N_j', the total number p of the variables is the only condition:

$$\text{for } M_i = N_j' \ldots i+j = 2^p - 1 \tag{11}$$

RULE IV

An incomplete set of variables at the input of an OR gate is similar to the situation of Rule II. Insertion of ZERO and ONE instead of missing or irrelevant variables is required to obtain pairs of maxterms, which represent the generated function as follows:

$$M_{i1}M_{i2}M_{i3}M_{i4} = N_{j1}'N_{j2}'N_{j3}'N_{j4}' \tag{12}$$

$$N_j' = \mathbf{f}_j'\mathbf{n}' = \mathbf{f}_j\mathbf{n}$$

$$\mathbf{f}_j = (1-\mathbf{f}_j') = (1-\mathbf{f}_i)^{-1} \tag{13}$$

The complemented minterm is a single term with respect to the complemented minterm matrix, n'. Expressions in minterms involve a function matrix f_j which is directly related to the original suffixes of the maxterms specified in f_i. If the matrix f_i is written in reverse order, the previous coefficient of the column 0 is placed in the last column to the right, proceeding with adjacent coefficients towards the left. This new "reversed" matrix is denoted by the exponent^{-1}. The following rules are obtained immediately:

$$\mathbf{f}_j = (1-\mathbf{f}_i)^{-1} = (1-\mathbf{f}_i^{-1}) \tag{14}$$

$$\mathbf{f} = \mathbf{f}_{j1}\mathbf{f}_{j2}\mathbf{f}_{j3}\mathbf{f}_{j4} = 1-(\mathbf{f}_{j1}'+\mathbf{f}_{j2}'+\mathbf{f}_{j3}'+\mathbf{f}_{j4}') \tag{15}$$

$$\mathbf{f} = 1-(\mathbf{f}_{i1}+\mathbf{f}_{i2}+\mathbf{f}_{i3}+\mathbf{f}_{i4})^{-1} \tag{16}$$

Equations 15 and 16 are applicable to express in function matrices incomplete sets of variables at OR gates.

Network Analysis

A logical system is defined by a set of blocks and the interconnections among these blocks. (Each block is itself a logical network.) Blocks with independent inputs only are designated as first-level blocks. Their outputs can be processed in another group of blocks, which belong to the second level. Eventually other blocks of higher level produce the final output.

The number of levels is pertinent to the classification of a system with regard to complexity, pulse distortion, and attenuation.[8]

One output Y_1 of a first-level block A_1 is defined by

$$Y_1 = \mathbf{y}_1\mathbf{n} = \mathbf{a}_1\mathbf{n} \tag{17A}$$

If this output Y_1 is processed in a block a_k of higher-level k, the output Y_k is obtained.

$$Y_k = \mathbf{y}_k\mathbf{n} = \mathbf{a}_k\mathbf{n} = \mathbf{f}(\mathbf{a}_1, \mathbf{a}_2 \ldots \mathbf{a}_k)\mathbf{n} \tag{17B}$$

All intermediate and final outputs Y_k are described by function matrix operators of the minterm matrix n, that is common to the whole system. It should be emphasized that the dummy indexes are identical with those of corresponding minterms. This reference to the states of the variables is maintained even if the minterm matrices are cancelled in all equations. Hence any matrix arithmetic can be restricted to the function matrices only to determine the relation f of all blocks involved.

$$\mathbf{y}_k = \mathbf{f}(\mathbf{a}_1, \mathbf{a}_2 \ldots \mathbf{a}_k) \tag{18}$$

Conclusions

This paper primarily concerns the introduction of function matrices and operators for the analysis of logical systems. In a subsequent paper, it is planned to expand this concept into a calculus capable of analyzing sequential operations.

It is shown in the appendix that this matrix algebra can be an economical means for analysis of logical networks of higher complexity without the aid of computers.

Appendix. Analysis of a 4-Level System with Eight Variables

In the control of a digital system following requirements expressed in conventional Boolean equations must be met.

$$Y_2 = (X_2+X_3'+X_5'+X_6+X_7+X_8')(X_1X_3'X_4X_6'X_7) \tag{19A}$$

$$Y_4 = X_1'X_2X_3X_4'X_5X_7Y_2' \tag{19B}$$

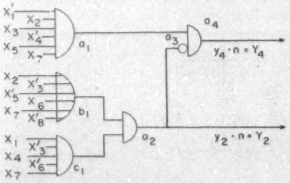


Fig. 5. Routine 4-level design of a network

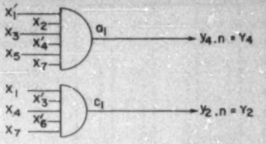


Fig. 6. Reduced 1-level network with identical function

Routine design leads to the network shown in Fig. 5. In view of the number of variables (eight) no attempt would have been made to simplify the circuit using conventional techniques such as the Havard chart.

In matrix notation the same problem is described as

$$\mathbf{y}_2 = \mathbf{b}_1\mathbf{c}_1 \tag{20A}$$

$$\mathbf{y}_4 = \mathbf{y}_2'\mathbf{a}_1 \tag{20B}$$

Or in applying the relation, equation 10,

$$\mathbf{y}_4 = (1-\mathbf{y}_2)\mathbf{a}_1 = (1-\mathbf{b}_1\mathbf{c}_1)\mathbf{a}_1 \tag{20C}$$

AND gate a_1 processes the inputs X_1' X_2 X_3 X_4' X_5 X_7, or in binary notation

01101√1√

Because two variables are missing, 2^2 minterms will be required and are defined by inserting ZERO and ONE in all combinations for the marked bits:

01101010(b)	106(d)
01101011	107
01101110	110
01101111	111

The second column is the decimal notation of the dummy index for the minterms or ONE's occurring in the function matrix a_1, corresponding to equation 21(A).

For the OR gate b_1 the appropriate dummy indexes i of the maxterms B_i are calculated and used to denote the ZERO coefficients of the appropriate minterms B_j in equation 21(B), with the understanding that all other matrix coefficients must be ONE in accordance with equation 16. It should be noted that j is simply the complement of i as shown in equation 11.

i = √10√0110
j = √01√1001

j_1 = 00101001(b)	41(d)
j_2 = 00111001	57
j_3 = 10101001	169
j_4 = 10111001	185

Five variables are present at the AND gate c_1; accordingly 2^3 or eight minterms will be generated as indicated in equation 21(C) and are obtained by the method used before for block a_1.

Because both outputs Y_2 and Y_4 are obtained from AND gates a_2 and a_4, respectively, only the 16 columns identified in the foregoing out of the total 256 columns need to be considered. As a matter of convenience the dummy index is written vertically and applies for all matrices; ZEROS are inserted only where pertinent.

Inspection of the function matrices

$$\text{dummy index}\left\{\begin{matrix}0 & & 0\;0\;0\;0\;0\;0\;1\;1\;1\;1\;1\;1\;2\;2\;2\;2 & & 2\\ 0 & & 4\;5\;7\;7\;8\;9\;0\;0\;1\;1\;6\;8\;0\;0\;1\;1 & & 5\\ 0 & \ldots & 1\;7\;3\;5\;9\;1\;6\;7\;0\;1\;9\;5\;1\;3\;7\;9 & \ldots & 5\end{matrix}\right\}$$

$$\mathbf{a}_1 = (\qquad\qquad 1\;1\;1\;1 \qquad\qquad) \tag{21A}$$

$$\mathbf{b}_1 = (1\ldots 1\;0\;0\;1\;1\;1\;1\;1\;1\;1\;1\;0\;0\;1\;1\;1\;1\;1 \ldots 1) \tag{21B}$$

$$\mathbf{c}_1 = (\qquad 1\;1\;1\;1 \qquad\qquad 1\;1\;1\;1 \qquad) \tag{21C}$$

$$\mathbf{a}_2 = \mathbf{b}_1\mathbf{c}_1 = (\qquad 1\;1\;1\;1 \qquad\qquad 1\;1\;1\;1 \qquad) = y_3 \tag{21D}$$

$$\mathbf{a}_3 = 1 - \mathbf{a}_2 = (1\ldots 1\;1\;0\;0\;0\;0\;1\;1\;1\;1\;1\;1\;0\;0\;0\;0\;1 \ldots 1) \tag{21E}$$

$$\mathbf{a}_4 = \mathbf{a}_1\mathbf{a}_3 = (\qquad\qquad 1\;1\;1\;1 \qquad\qquad) = y_4 \tag{21F}$$

directly reveals that the following conditions exist:

$$\left\|\begin{matrix}y_4\\ y_3\end{matrix}\right\| = \left\|\begin{matrix}a_1\\ c_1\end{matrix}\right\| \tag{22}$$

The blocks b_1, a_2, a_3, and a_4 do not contribute to the outputs in any respect and are redundant.

Fig. 6 indicates the reduced network involving a single level which performs the same function. The analysis eliminates four of six blocks.

References

1. Giant Brains or Machines That Think (book), E. C. Berkley. John Wiley & Sons, Inc., New York, N. Y., 1949.

2. Mechanized Reasoning, D. M. McCallum et al. *Electronic Engineering*, London, England, Apr. 1951, pp. 126–33.

3. Feedback Logical Computers, D. M. McCallum et al. *Ibid.*, Dec. 1951, pp. 458–61.

4. Machine Aid for Switching Circuit Design, C. E. Shannon, E. F. Moore. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Oct. 1953, p. 1348.

5. The Problem of Simplifying Truth Functions, W. V. Quine. *American Mathematical Monthly*, Menasha, Wis., vol. 59, Oct. 1952, pp. 521 ff.

6. Elements of Boolean Algebra for the Study of Information Handling Systems, R. Serrell. *Proceedings*, Institute of Radio Engineers, vol. 41, Oct. 1953, pp. 1366–80 (14 entries).

7. Tensor Analysis of Networks (book), G. Kron. John Wiley & Sons, Inc., 1949.

8. Engineering Multistage Diode Logic Circuits, B. J. Yokelson, W. Ulrich. *AIEE Transactions*, vol 74, pt. I, Sept. 1955, pp. 466–75.

The Economics of Component Parts Selection

E. F. FUEGEL
MEMBER AIEE

IT IS OFTEN SAID that the best buy is not necessarily the lowest cost article which can be purchased. It is usually the article which will perform the required function for the required time with minimum over-all cost. Equipment may be designed to meet one or intermediate grades of the following requirements:

1. Lowest manufacturing cost (where maintenance costs are unimportant).
2. Minimum total cost (consisting of manufacturing, installation, and maintenance costs for the required time the equipment is to be used).
3. A minimum number of failures for the required time. (No compromises are to be made in manufacturing, installation, and preventive maintenance to achieve this goal.)

The first requirement involves the selection of parts for which the principal criteria are minimum purchase price and installation cost. Items such as toys are in this category. The economics involved for this requirement are obvious and will not be discussed in this paper. The third requirement, in general, involves customers concerned with the protection or maintenance of human life, where the best available components are selected and economics are trivial.

Most industrial equipment or systems must meet requirement 2 or be intermediate between requirements 2 and 3. By contrast, entertainment and domestic equipment generally is in the range of requirements 1 and 2. Because a good share of the total electronics business falls in this category, a large number of component parts—tubes, transistors, resistors, capacitors, and transformers—are designed specifically for requirements 1 and 2.

When such components are applied to systems that must meet requirements 2 and 3, as in the case of many business machines, the result is often unsatisfactory performance. To design a satisfactory system, i.e., one that meets the necessary levels of reliability, the engineering team must possess numerous abilities, including a thorough knowledge of the requirements, environmental conditions, design technique, and component parts selection and application.

Reliability

Business machines, like many types of electronic equipment, are designed to provide the maximum output for minimum cost over a long period of time. A further requirement is that erroneous output information cannot be tolerated. The equipment should cease to function, or the choice to continue or stop processing information should be under the control of the programmer, or the equipment should be self-correcting, with an indication of what failed when maintenance is required. Where pay checks are involved, for example, the cost of errors can become intolerable.

The reliability of the system must be high. Reliability is defined as the probability that a system, when operated as intended, will provide the desired performance for the desired period of time. If a component part is used in significant quantities, the probability that a certain percentage of these parts will be within certain limits under known operating conditions can be determined, when the rating, parameter limits, environmental conditions, and time are defined.

For purposes of this discussion, the following definitions are assumed:

1. *Component part.* Basic unit, such as tube, transistor, resistor, capacitor.
2. *Component.* Grouping a number of component parts into a circuit, for example, an oscillator, register, or counter.
3. *Equipment.* A number of components, usually housed in a cabinet, making up units like a card punch, sorter, tape storage unit, or memory. An equipment may be a complete entity (card punch, for example) or it may combine with other equipments to form a larger system.

Paper 57-678, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted August 7, 1956; made available for printing July 9, 1957.

E. F. Fuegel is with the International Business Machines Corporation, Poughkeepsie, N. Y.

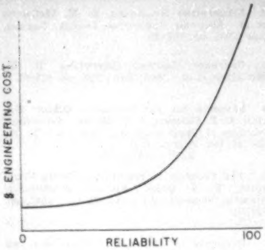
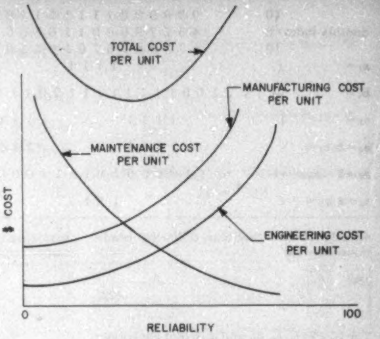


Fig. 1 (left). Diagram showing how engineering costs increase as desired reliability increases

Fig. 3 (right). Illustration of how higher manufacturing costs are sometimes necessary to obtain the desired reliability



4. *System.* A complete operating unit, such as a digital computer, generally made up of a combination of equipments.

Duplicate Systems

The user of a system or equipment might define reliability as "the probability that the system or equipment will do satisfactorily what it is expected to do when I want it done." One way by which the customer can obtain greater reliability is to buy two machines. Then if one does not work, he can use the other one. The odds on completing the job with two machines are better than with a single one. Whether the second machine can be called redundant depends on two factors: the reliability of the first machine, and the reliability that the customer requires. If machine reliability is 95% and the customer is satisfied with 75%, the second machine is redundant. If the best machine available works only 50% of the time, then the customer must have two to meet his requirements. Duplicating entire machines is an extreme measure. The cost is generally prohibitive unless the second machine can be used for supplementary operations that do not have to be performed at specific times.

Duplicate Components

Another possible way to improve reliability is to duplicate unreliable components. This is only possible when the additional circuitry and parts required do not lower the over-all reliability of the system. In general, duplicating selected components rather than entire systems is preferable from a standpoint of cost, size, and weight. The duplication procedure may be expanded to three or more parallel sets of components, as required to obtain the necessary reliability.

The over-all reliability gained by duplicating components may be greater than that obtained from duplicate systems. The use of duplicate systems versus duplicate components may be considered analogous to two chains holding a load; if one link breaks, any link in the second chain must be able to hold the load. If a single chain is made with the same links so that two links are in parallel throughout the length of the chain, the failure of one link will mean that only one single link will be required to hold the load. Intuitively, it can be reasoned that the second chain is more reliable than the pair of chains.

The use of duplicate rather than single components is generally not advantageous from the viewpoints of cost, size, and weight. It should only be resorted to after all the other ways of obtaining the necessary reliability have been thoroughly explored. These include improved manufacturing and quality control methods, better design, better application of component parts, the use of duplicate component parts, and the use of marginal checking.

Marginal Checking

This is a method for detecting progressive degradation before actual equipment failures occur. It can be done as part of a preventive maintenance procedure, to improve system reliability from the customer's point of view. However, this type of checking does not spot potential catastrophic failures. Also, rapid marginal checking of the component parts

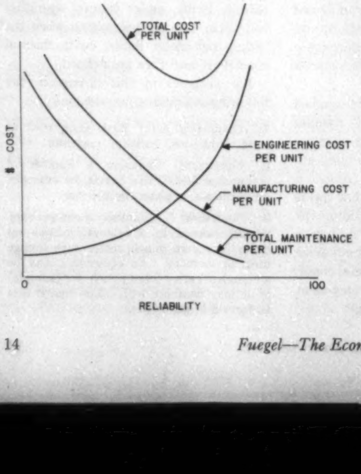
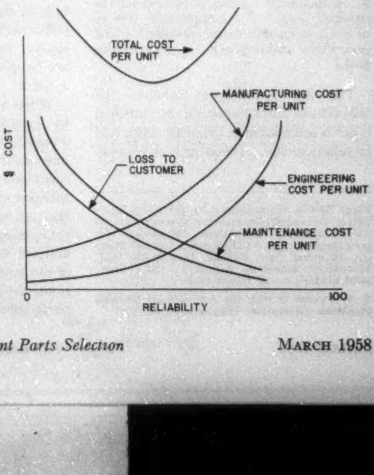


Fig. 2 (left). Curves showing that an increase in engineering effort can improve reliability and decrease manufacturing costs

Fig. 4 (right). Diagram including "loss to customer" cost which can affect the total cost curve



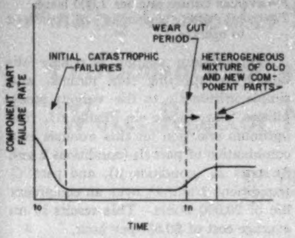


Fig. 5. A typical curve for the failure rate of component parts versus time

in equipment may require additional circuitry that will increase the total size, weight, and cost. The additional cost of such equipment, the cost of maintenance and periodic testing, the reliability obtained, and the availability of the equipment to the maintenance man are factors that must be considered carefully before a marginal checking system is advocated.

Rating of Component Parts

It is well known that many good component parts are available. A large percentage of equipment failures are caused by poor selection, control, and application of component parts. To design reliable equipment, the circuit designer must apply these components in a way that provides the best performance.

Using the manufacturer's rating or arbitrarily altering it to one half or twice the rating will not necessarily provide the desired results. The probable life of a component part depends on how it is applied. The manufacturer's rating is often arbitrary and is usually not the best rating for the equipment which is being designed. Frequently it is possible to select less expensive component parts and obtain the desired reliability by derating.

The tolerance on the purchased component part is the manufacturing tolerance. It in no way indicates how much larger the tolerance will be after the part has been in service for the desired time. It is this larger tolerance that is applicable and it will be referred to here as the design tolerance

The rating and the design tolerances of component parts which can be manufactured within reproducible limits should be based on the period of time during which the equipment is expected to operate. The rating and design tolerance which is determined by life-testing are based on the probability of a certain percentage of component parts remaining within the design tolerance for the required time. The basis for the rating, design tolerance, and survival percentage includes the availability and the cost of the parts.

The life testing of component parts to obtain the probable survival percentages for intermediate times is expensive but necessary. Manufacturers of component parts generally do not have data available for each type of equipment requirement because of the immense variety of requirements and the expense of obtaining the data. Many manufacturers only test their own and competitors' products to determine whether they have improved their product, and to see what their competitive position is.

Table I. Selection and Application of Typical Component Parts

Component Part	Purchase Cost of Part (P)	Labor Cost to Replace (L_o)	Number of Parts (N)	Average Failure Rate per 1,000 Hours (F) for T_n			
				T_n 5,000	T_n 10,000	T_n 20,000	T_n 50,000
A_1	$0.80	$5.00	60	0.01	0.01	0.01	0.01
A_2	1.90	3.00	60	0.005	0.005	0.005	0.005
A_3	3.50	3.00	60	0.001	0.001	0.001	0.001
B_1	0.03	8.50	360	0.0005	0.001	0.002	0.005
B_2	0.15	5.00	360	0.001	0.0006	0.0008	0.001
B_3	0.50	5.00	360	0.0001	0.0001	0.0002	0.0003
C_1	0.25	4.50	250	0.001	0.002	0.003	0.004
C_2	0.75	4.50	250	0.0001	0.0001	0.00015	0.0002
D	Large number of miscellaneous parts; maintenance estimated from experience on previous equipments.						

Conditions:

1. If quantity of part *B* is increased by 20%, failure rates of *A* decrease by 10%.
2. If design tolerance of *B* is reduced by 3%, failure rates increase: B_1 by 25%; B_2, 10%; B_3, 5%. However, quantity of part *A* is reduced 15% and the failure rates remain the same. Also, population of part *C* is reduced 15%.
3. If the voltage rating of part *C* is halved, the cost of all parts *C* is reduced 40%. However, failure rate of *C* increases by a factor of 8.
4. Ten per cent of the part *C* failures will cause an equal number of additional parts *A* to fail.
5. Loss to the customer (L_c) each time a failure occurs is $6.00.

Why Life-Test Component Parts?

The advantages of life-testing component parts are as follows:

1. An estimate of the reliability of a component or system may be determined from the component part reliabilities before even a prototype is built. Changes can be made more readily in the design stage to obtain the desired reliability.
2. A basic group of component parts can be used in a large number of components. A number of samples must be life-tested to obtain reliability data, which in the case of components or systems may be more expensive, especially if a number of changes have to be made to obtain the desired reliability.
3. The time involved in life-testing components or systems is prohibitive when a life expectancy of 20,000 or 50,000 hours is desired. Utilizing standard component parts facilitates reliability and reduces the time and over-all cost. The component parts art does not change as rapidly as the requirements for new circuits. Therefore, estimates of the reliability of parts can be obtained before parts are used in circuits.
4. Improvements and cost reductions in component parts can be recognized and incorporated in existing or new components if a continued surveillance is made on component part reliability.

It is not to be inferred that evaluation of the component or system is not necessary. The component must be evaluated to determine that the component parts are applied properly, including the dependence of component parts on each other, and that the component will function as desired initially, and for the desired life expectancy.

Cost Considerations

Engineering Costs

The cost of engineering a reliable product usually increases as the desired reliability increases. This relationship is shown in Fig. 1. It is usually more expensive to achieve a simple design than a complex one. Designing equipment that will minimize manufacturing errors takes considerably more effort than evolving designs which do not take manufacturing difficulties into account. As reliability is increased, the costs of evaluation, testing, design, and control all increase.

Total Cost

The manufacturing cost may either increase or decrease as a function of reliability, depending on the type of equipment, manufacturing methods, manufacturing difficulties, design complexity, and the selection and control of materials

Table II. Cost of Parts Based on Data from Table I

Component Part	Applicable Conditions	Equipment Life in Hours (T_n) 5,000	10,000	20,000	50,000
A_1	2	$ 72			
A_1	2		$106		
A_2	2, 3			$157	
A_2	2				$213
B_1	2	27			
B_1	2		76		
B_2	2			125	
B_2	2				245
C_1	2	64			
C_1	2		99		
C_2	2, 3			151	
C_2	2				183
D		100	150	300	$1,900
Total cost		$263	$431	$733	$2,541
Cost per hour		$0.053	$0.043	$0.037	$0.051
Average no. of failures excluding (D)		4.9	14.2	17.1	10.6

in the product. It is often possible to improve reliability and reduce manufacturing costs by increasing the engineering expenditure. Such a situation is illustrated in Fig. 2. With some equipment, however, the desired reliability can only be obtained by complex manufacturing procedures and controls, more expensive materials, and complex designs. In such cases the manufacturing cost will increase as reliability increases, as shown in Fig. 3.

The sum of the manufacturing, maintenance, and engineering costs over the desired life is the criterion that determines the optimum allocation of expenditures. Thus, in Figs. 2 and 3 the total cost per unit curve is the most significant one, and the lowest point on it is the ideal operating point, economically speaking. However, in most cases customer demands for high reliability must be considered, and compromise made on the optimum operating point. From the plant controller's point of view, the figures can be used in determining the ratio of income to expenditures.

CUSTOMER LOSSES

Another factor that should be considered in the total cost is the loss to the customer caused by equipment failure (as shown in Fig. 4). If other costs remain the same, the "loss to customer" cost will have the effect of shifting the total cost curve towards higher reliability. The required reliability is that which will provide the maximum profit to the customer. Optimum reliability is usually determined by past experience and good judgment, taking into account the factors previously discussed plus items such as competitive position, time, state of the art, taxes, interest, and market conditions. As more information becomes available by increased use of computers and accounting systems, it will be possible to determine optimum reliability more accurately. It may be possible to use methods involving successive approximations.

Example

Experience with a number of equipments indicates that the failure rate versus time characteristic for large numbers of a component part type generally takes the form shown in Fig. 5, where parts are replaced in the system immediately after a failure occurs. The plateau at the right in Fig. 5 is the result of replacing failed component parts, since these replaced parts generally have a higher probability of surviving to a specified time than those which have been in the system longer. The average failure rate is the portion of the population which fails during a specified time.

Tables I and II illustrate the economical selection and application of parts, using the principles previously described. Table I lists groups of typical component parts, showing numbers used, purchase and replacement costs, and failure rates. (These are average failure rates per 1,000 hours for periods of 5,000, 10,000, 20,000, and 50,000 hours. They depend on how the component parts are applied in the circuits, and on the environmental conditions. E. E. Brewer[1] has discussed various methods of obtaining operating failure rates from component part failure rate data.) Also shown in this table are five assumed conditions that will govern the calculations of this example.

Table II shows the costs as worked out from the data of Table I, and indicates the optimum selection of parts *A*, *B*, and *C* under conditions 1, 2, and 3 (conditions 4 and 5 apply to all selections). The cost of a part, neglecting factors like interest, installation cost, and inspection cost, is calculated from the following formula:

$$P(N)+(P+L_a+L_o)(N)(F)(T_n/1,000)$$

where

P = purchase cost of part
N = number of parts
L_a = cost of replacing failed part
L_o = loss to customer when part fails
F = average failure rate per 1,000 hours
T_n = total time equipment is expected to operate

The optimum solution is readily obtained by applying this formula and simple arithmetic to the various possibilities. As shown in Table II, the optimum selection for this example is a combination of part A_2 (conditions 2 and 3), part B_2 (condition 2), and part C_2 (conditions 3 and 2), over an equipment life of 20,000 hours. This results in an average cost of $0.037 per hour.

Conclusions

To make improvements in the reliability evaluations and procedures, it is necessary to have a continuous field survey program that correlates laboratory and design procedures with actual field equipment and component part failures. Also, a constant watch must be kept on the quality control inspection of incoming component parts, to see if the part manufacturer is providing a reproducible product within the specified limits. This is necessary because the probable tolerance limits after a specified period of time depend on the probable limits at time zero and the probable aging characteristics of the parts.

The first step in arriving at the best equipment for a particular requirement is a thorough knowledge of the economics, selection, application, and performance of component parts. Thereafter, reliability may be improved in several ways. Expending more engineering effort may make it possible to use component parts with broader design tolerances. This generally will decrease the purchase cost of the parts and improve the reliability of the equipment. Reducing the number of component parts may increase reliability and at the same time cut manufacturing costs. Higher reliability may be achieved by increasing the number of components parts to allow the use of lower cost parts with broader tolerances. Likewise, the most economical results may be realized by duplicating components or using periodic maintenance marginal testing methods.

Reference

1. MATHEMATICAL TOOLS FOR RELIABLE DESIGN. E. E. Brewer. *Proceedings*, Symposium on Reliable Application of Electron Tubes, Radio, Electronics-Television Manufacturers Association. New York, N. Y., May 21–22, 1956.

Analog Computer Study of the Transient Performance of a Dual-Cycle Boiling-Water-Reactor Nuclear Power Plant

DAVID W. LEIBY
ASSOCIATE MEMBER AIEE

AS THE COMPLEXITY of multiloop systems increases, the prediction of their transient performance with the use of conventional linear analytical methods becomes quite difficult. In those cases in which each loop of the system contains several nonlinearities, the application of analog computer techniques becomes an attractive method in the solution of transient performance problems. This paper presents a résumé of an analog computer study of the transient performance of such a complex nonlinear system as represented by a dual-cycle boiling-water-reactor nuclear power plant.

As the first step in studying the power plant transient performance, the system was broken down into a number of individual closed loops. These were simulated and analyzed on the analog computer; then, the total system was interconnected and over-all system performance obtained. In the simulation of the individual control loops and in their interconnection, various simplifying assumptions were made. These assumptions were necessary both to limit the complexity of the simulation to the available simulation equipment, and also to cover areas in which detailed information was not immediately available on system phenomena and parameters. From the results obtained, predictions may be made concerning the general trends in the performance of the power plant for various system disturbances, and the importance of system parameters and their effects upon system performance. This information is quite useful in the determination of design compromises in which certain areas of performance must be sacrificed for considerations in other areas.

Description of Boiling-Water-Reactor Power Plant

The Dresden nuclear power plant is being designed to operate on the dual-cycle boiling-water-reactor principle, which has been described in detail in previous articles in the literature.[1,2] Basically, a portion of the energy is withdrawn from the reactor in the form of high-pressure steam obtained by direct boiling within the reactor core, and a portion of the output is taken in the form of a lower-pressure steam obtained from a heat exchanger. A simplified block diagram of an early version of the Dresden nuclear power plant is shown in Fig. 1, which gives the main components of the power plant system as well as the interconnections of the various flow paths of water and steam. The primary recirculating loop consists of the reactor vessel with its reactor core, the risers through which the mixture of water and steam goes to the high-pressure steam drum and the steam separators, the primary side of the heat exchanger, and the interconnecting recirculating water lines. The high-pressure steam flows either through the bypass valve directly to the condenser or through the primary steam valve to the high-pressure stage of the turbine. The secondary steam generated in the secondary side of the heat exchanger flows through the secondary steam valve to an intermediate pressure stage of the turbine. The condensate from the turbine passes through the associated feedwater equipment and then returns proportionally through the feedwater pumps to the secondary steam drum and into the recirculating water loop to maintain proper water level within the various components of the system. The multitude of auxiliary equipment and functions such as blow-down, demineralizers, make-up water, etc., have been omitted from the block diagram, since their magnitudes and operations are assumed to have insignificant effects upon the transient performance of the power plant during normal operation.

The valves associated with the high-pressure steam flow, the bypass flow, and the secondary steam flow are interconnected through the mechanical and hydraulic system illustrated in Fig. 2. The error signal between the reactor vessel pressure and the reference pressure operates through a pressure regulator to position the mid-point of the pressure link. The error signal between the turbine speed and the reference speed acts through a speed governor to position the mid-point of the speed link. The secondary admission valve is connected through a hydraulic amplifier to the right-hand end of the speed link, the primary steam admission valve is connected through a hydraulic amplifier to the interconnection between the speed link and the pressure link, and the bypass valve is connected through hydraulic amplifying equipment to the left end of the pressure link. In addition, the left-hand end of the pressure link is restrained by a spring to hold the bypass valve in the closed position.

This unique linkage system provides automatic control of both the turbine speed and the reactor vessel pressure when the power plant is operating into an isolated load. If the load is increased upon the turbine, the speed will tend to droop, causing the speed governor to raise the mid-point of the speed link, thereby opening the secondary steam valve to maintain the proper speed on the turbine. If the reactor vessel pressure tends to increase over the reference value, the pressure regulator will raise the mid-point of the pressure linkage, causing the pressure link to rotate about its left end as a fulcrum and the speed link to rotate about its mid-point, thereby opening the primary steam valve and closing the secondary admission valve. If there should be more primary steam flow available than required to maintain load on the turbine, the secondary admission valve will become seated and any further demand for more high-pressure steam flow will cause the output of the pressure regulator to override the restraining spring, thereby opening the bypass valve to the condenser.

One of the attractive characteristics of a boiling-water reactor is its stable operation at constant power when it is operating under constant pressure.[3,4] When the dual-cycle feature is applied to a boiling reactor it is found that the reactor power level is a function of the subcooling of the water entering the bottom of the core. The term "subcooling" is used to describe the difference in enthalpy between the water entering the bottom of the core and the saturation enthalpy at

---

Paper 57-684, recommended by the AIEE Computing Devices and Nucleonics Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 22, 1957; made available for printing July 18, 1957.

DAVID W. LEIBY is with the General Electric Company, Schenectady, N. Y.

This paper is based on a study performed for the Atomic Power Equipment Department of the General Electric Company.

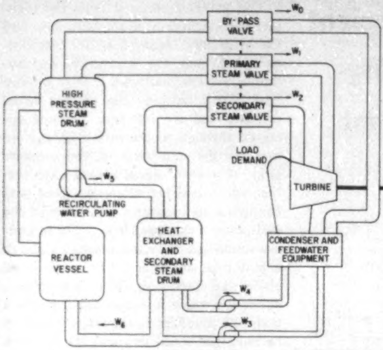


Fig. 1. Dual-cycle boiling-reactor power plant

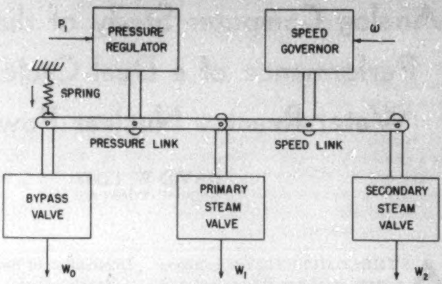


Fig. 2. Steam flow control

the operating pressure. This subcooling of the recirculating water is essentially a measure of the amount of energy being extracted in the secondary steam system plus the energy required to heat the primary steam flow condensate up to the boiling point.

The manner in which the subcooling is related to the total power output of the reactor may be seen by referring to Fig. 3, where three functions within the reactor core are indicated relative to their positions along the axis of the reactor for two different power levels. If the heat flux density from the fuel elements to the coolant steam $Q_\phi(y)$, is assumed to be sinusoidally distributed along the axis of the core, the energy added to the coolant stream, $\int Q(y)$, has the form of a $(1-\cos\theta)$ curve. At some point upward through the core, designated as α, sufficient energy has been added to the coolant stream to bring the recirculating water up to the boiling point. All energy added to the coolant stream from that point upward causes boiling within the reactor core. The difference in the abscissa between the total energy added curve and the subcooling value may be considered as an indication of the bubble concentration. The total area beyond the subcooling boundary line gives an indication of the total steam volume within the reactor core. The steam bubbles within the core cause negative reactivity which must be counterbalanced by positive reactivity from the control rods. For the purpose of this study, the negative reactivity from the steam bubbles is considered to be proportional to the integral of the local bubble population times the relative flux level squared. The reactivity from the bubbles or voids, as a function of distance through the reactor, is given by the $\Delta K_v(y)$ curve with the total reactivity from the voids being represented by the area under this curve.

The two similar diagrams in Fig. 3 are shown for operating at one-half rated power level and for operation at full rated power. In the one-half power condition, shown in Fig. 3(A), the subcooling is equal to approximately 20% of the full rated output of the reactor, causing boiling to begin approximately 40% through the reactor. It may be seen that if the subcooling is increased, the bubble population will diminish, thereby causing an effective decrease in the area representing the reactivity from the voids. Since this reactivity term is negative a decrease in its magnitude puts an effective positive reactivity into the reactor, thereby causing an increase in the reactor power.

When the reactor is operating at full power as shown in Fig.3(B), boiling does not begin until the recirculating water has penetrated over 50% through the reactor. This gives a much shorter region in which the bubbles are present but, due to the increased magnitude of the bubble concentration, the area of the curve which represents the total reactivity of the voids is the same as that in the conditions shown in (A). Since the subcooling of the water entering the bottom of the reactor core Q_s is essentially proportional to the energy being extracted in the secondary steam system, a method of control of the reactor power level over a reasonable power level range is obtained by changing the rate at which energy is withdrawn from the secondary steam system.

Power Plant Simulation

The presentation of all the details of the derivation of system transfer functions and simulations of all components of the power plant is well beyond the scope of this paper. Therefore only the block diagrams of the individual loops within the system will be presented along with a discussion of several of the simulation problems involved in each loop.

REACTOR-VOID FRACTION LOOP

Probably the most important single control loop within the Dresden nuclear power plant is that which expresses the relationship of the reactor power level, the recirculating water subcooling, and the effective reactivity from the steam voids. This loop, which is present within the reactor core, may be represented by a 5-element block diagram as shown in Fig. 4. The reactor kinetics represents the relationship between the reactivity Δk, applied to the nuclear reactor and the output of neutron flux Φ. The fuel element thermal model gives the time-dependent behavior of the heat energy output at the surface of the fuel elements Q_ϕ as a function of the neutron flux level as an input. The void fraction model expresses the relationship between the reactor heat flux and the recirculating water subcooling Q_s, to give an effective steady-state reactivity from the voids Δk_{v0}. The void generation time simulation block considers the time for the growth of the bubbles within the reactor core and the accumulation of the effect of bubbles as a function of the transport time through the core. The fifth block represents the pressure dependency of the effective reactivity in that, as the reactor vessel pressure P_1 increases, a given amount of energy in steam will represent a smaller volume of voids within the

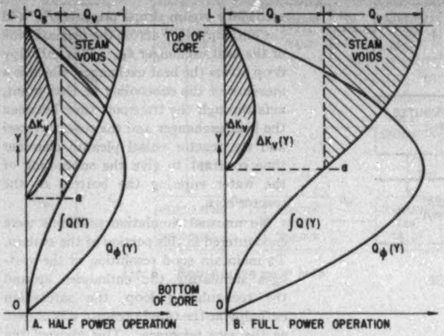


Fig. 3. Reactor void-fraction relationships

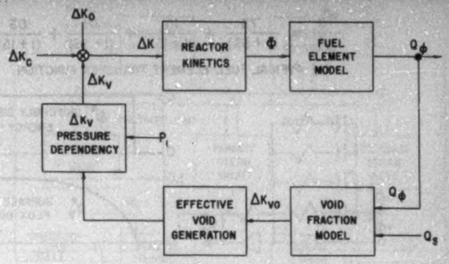


Fig. 4. Reactor void-fraction loop

reactor core. The reactivity from the voids is added with the proper polarity to the reactivity from the control rods, temperature coefficient, Doppler effect, poisons, etc., to give an effective ΔK of zero under steady-state operating conditions. The reactor kinetics is simulated in a conventional manner, which has been described at various places in the literature, by the use of one operational amplifier with the five delayed neutron groups connected across it and one electronic multiplier to obtain the power level dependency of the reactor gain factor.[5,6]

During the early phases of the study of the Dresden nuclear power plant, the fuel element transfer function was considered to be a simple time constant. During the investigation of the stability of the reactor-void fraction loop, a more refined model of the fuel element was found desirable. An expression was derived for the heat transfer to the coolant at the surface of an infinite cylinder having uniform heat generation throughout its interior. Further improvement of the thermal model of the fuel element was made by considering the thermal impedance present in the bond between the fuel material and the jacket of the fuel element.[7] The transfer function for the heat flow across the boundary between the fuel element and the coolant may be expressed as a series of single time constants which may be readily simulated on the analog computer.

The gain constants and time constants in the series expression are functions of the Eigen values of the Bessel functions, the outer radius of the fuel element, the thermal diffusivity of the fuel element material, and the thermal resistance between the fuel element and the jacket. The transfer function of a typical fuel element is shown in Fig. 5, along with its analog computer simulation and the transient response of the fuel element to a step change in uniform heat generation within the fuel element.

The transfer function of the typical fuel element indicates that 75% of the energy occurs through a time constant of 5 seconds, 15% through a time constant of 1 second, and 5% through a time constant of 1/2 second, with the remainder of the energy transfer occurring in smaller and smaller parcels through faster and faster time constants. The response of such a transfer function to a step input then shows up with an infinite slope at the beginning of the transient which decreases as the effect of the faster time constants disappear and the remaining energy comes from the longer time constant contribution.

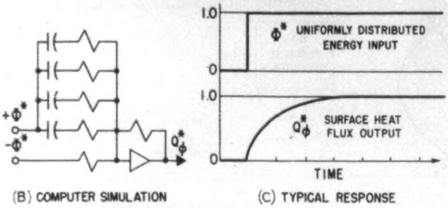
In the simulation of the void fraction model it was necessary to reproduce the calculated effect of the subcooling of the water entering the bottom of the reactor core Q_s, and the operating level of the reactor Q_ϕ, upon the effective reactivity obtained from the voids. By means of the geometry presented in the void fraction relationships shown in Fig. 3, an analytical expression may be derived for the effective reactivity of the voids as a function of these two system variables, giving the family of curves as shown in Fig. 6. In this figure, each curve represents a given constant power level, the ordinates indicate values of effective reactivity from the voids, and the abscissas represent values of subcooling entering the bottom of the reactor core. This family of curves, or void map as it is often called, is so set up in a per-unit system that for the unit power output curve, unit subcooling causes zero void generation, whereas zero subcooling causes unit effective reactivity from voids. In the power plant configuration under discussion, the operating line is such that full power is obtained at approximately 50% per-unit subcooling. Therefore, the operating line drawn on this family of curves shows that only about 20% of the effective ΔK from voids is present of that which would be present if the reactor were boiling completely at rated power with saturated water entering the bottom of the core. This family of curves shows that with the operating line at the indicated position, control from 100% power down to approximately 25% power is possible by means of variation of the subcooling.

The behavior of the reactor loop may readily be seen by considering operation about some point on the operating line, say at the 0.8 power position, where the subcooling is approximately 37%. If the subcooling were increased to 40%, the operating point would move down along the constant 80% power line which would cause a decrease in the effective reactivity from the voids, thereby causing the reactor power to increase. The reactor power level will then increase to some value of approximately 85% to bring the operating point back to the operating line. In a similar manner, if the subcooling were decreased down to 30% rated value, the operating point would move upward to cause an increase of about 30% in the reactivity from the voids, thereby reducing the reactor power to approximately 70% to bring the operating point back to the operating line.

This family of curves is simulated on the analog computer equipment,[8] as shown in Fig. 7. An electronic divider circuit is operated on by the subcooling Q_s, by the heat flux from the reactor Q_ϕ, and by a

$$\frac{Q^*_\phi}{\Phi^*} = \frac{.75}{(1+5S)} + \frac{.15}{(1+1S)} + \frac{.05}{(1+.5S)} + \frac{.03}{(1+.1S)} + \cdots$$

(A) TYPICAL FUEL ELEMENT TRANSFER FUNCTION



$+\Phi^*$
$-\Phi^*$
Q^*_ϕ
1.0
0
Φ^* UNIFORMLY DISTRIBUTED ENERGY INPUT
1.0
Q^*_ϕ SURFACE HEAT FLUX OUTPUT
0
TIME

(B) COMPUTER SIMULATION

(C) TYPICAL RESPONSE

Fig. 5. Fuel element thermal model

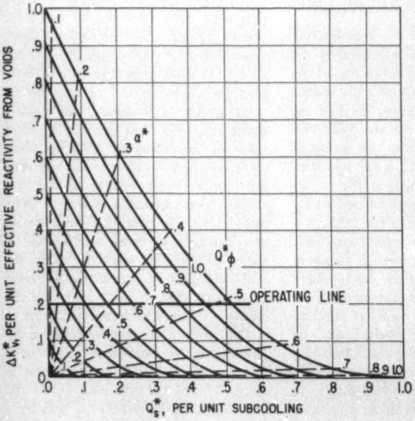
bias voltage. The output of this divider is a function of the reactivity from the voids. This function is then applied to a function generator which simulates the per-unit power curve of reactivity as a function of subcooling. This function generator uses the forward characteristic of silicon diodes and can reproduce the curve with less than 1.0% error. The simulation of the void fraction map was adjusted to maintain the spacing and the slope of the power curves near the operating line, allowing some deviation at the extremities of the curves. These deviations were not considered serious since normal operation did not cause violent operational departures from the operating line and also because the assumptions made in the calculations of this curve did not warrant the additional effort to obtain a more accurate representation.

RECIRCULATING WATER LOOP

The block diagram of the recirculating water loop, shown in Fig. 8, includes eight components. The thermal model of the reactor vessel takes the primary steam flow W_1, and the difference between the reactor heat rate Q_ϕ and the subcooling Q_s, as the input signals, and produces the pressure in the reactor vessel P_1 as the output signal. The reactor vessel pressure determines the enthalpy h_{60} of the recirculating water leaving the reactor vessel. This enthalpy acts through the time constant of the water storage within the primary steam drum and the transport time between the steam drum and the heat exchanger to give an enthalpy h_{62} at the input to the primary side of the heat exchanger. The steam flow from the secondary side of the heat exchanger W_2 produces the heat exchanger secondary steam drum pressure P_2, and the enthalpy drop across the primary side of the heat exchanger h_{63}. This enthalpy drop across the heat exchanger, which is a measure of the subcooling at that point, acts through the transport time between the heat exchanger and the reactor vessel and the reactor vessel plenum chamber time constant to give the subcooling of the water entering the bottom of the reactor core.

No unusual simulation problems were encountered in this portion of the system. To maintain good resolution of the voltages simulating the enthalpies around the recirculating loop, the saturation enthalpy at the rated pressure was chosen as the zero reference. This allowed the changes in the subcooling to cover the full range of output voltage of the operational amplifier. Each transport time was simulated by an all-pass network utilizing three operational amplifiers.[9] The heat exchanger was simulated by a relatively simple analysis of the single-pass, cross-flow, U-tube type of configuration.[10] A typical response of this portion of the system is given in Fig. 9, in which a step decrease in the secondary steam flow from rated value to zero is applied, such as would occur during a turbine tripout. The secondary steam pressure and the subcooling appearing at the bottom of the reactor are indicated as the output functions. The pressure within the secondary steam drum rises approximately as a single time constant whose magnitude is determined primarily by the weight of water stored within the secondary side of the heat exchanger. The subcooling leaving the heat exchanger follows approximately

Vol.
Pt. 1
1958



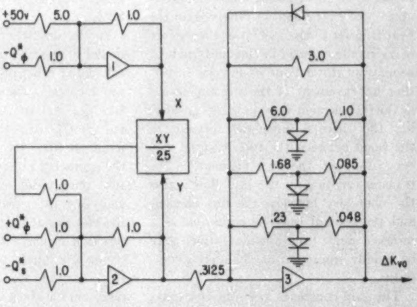


Fig. 7. Reactor void-map simulation

Fig. 6 (left). Calculated effective void map of ΔK_v^* as a function of reactor heat output Q_ϕ^*, and subcooling Q_s^*

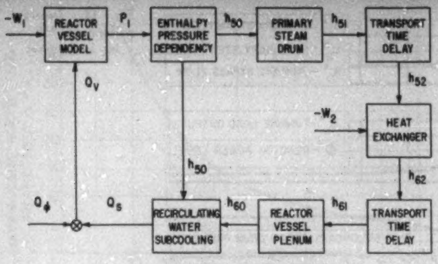

$-W_1$
REACTOR VESSEL MODEL
P_1
ENTHALPY PRESSURE DEPENDENCY
h_{50}
PRIMARY STEAM DRUM
h_{51}
TRANSPORT TIME DELAY
h_{52}
Q_V
$-W_2$
HEAT EXCHANGER
h_{50}
h_{62}
Q_ϕ
Q_S
RECIRCULATING WATER SUBCOOLING
h_{60}
REACTOR VESSEL PLENUM
h_{61}
TRANSPORT TIME DELAY

Fig. 8. Recirculating water loop

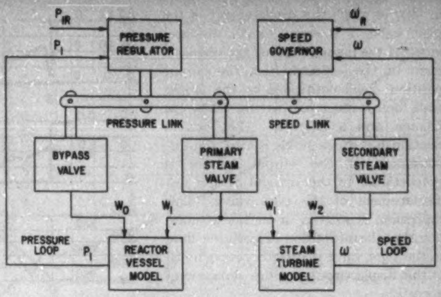

P_{IR}
P_1
PRESSURE REGULATOR
SPEED GOVERNOR
ω_R
ω
PRESSURE LINK
SPEED LINK
BYPASS VALVE
PRIMARY STEAM VALVE
SECONDARY STEAM VALVE
W_0
W_1
W_1
W_2
PRESSURE LOOP
P_1
REACTOR VESSEL MODEL
STEAM TURBINE MODEL
ω
SPEED LOOP

Fig. 10. Pressure and speed-loop interconnection

the same curve as the secondary steam drum pressure. It appears at the bottom of the reactor core after being operated upon by the transport time between the reactor vessel and the heat exchanger, and the time constant in the reactor vessel plenum.

PRESSURE LOOP AND SPEED LOOP

In the preliminary studies leading up to the total power plant simulation on the analog computer equipment, individual studies were made of the pressure loop and the speed loop. In these studies, the transient performance of the individual loops was investigated for a large number of variations in system parameters and for various driving functions. Since the pressure loop and the speed loop are quite closely interconnected, both through the cross-coupling of their control signals and also by an interconnection of their main integrator elements, it is more convenient to present a block diagram of the combined pressure and speed loops as shown in Fig. 10, rather than the individual block diagrams. The pressure loop under normal operating conditions consists of the pressure regulator, the primary steam valve, and the reactor vessel thermal model. The speed loop consists of the speed governor, the secondary steam valve, and the steam turbine model.

The steam flow control system contains many more individual elements than those shown in the pressure and speed loop interconnection block diagram of Fig. 10. In many places throughout the system, the power requirements are such that several stages of hydraulic amplification devices must be cascaded in order to obtain proper operation of the system. Also, the system includes auxiliary control equipment such as pre-emergency pressure regulators and special hydraulic mechanisms for obtaining particular operating characteristics of the system. To simulate properly the performance of the pressure and speed loops, it is necessary to include the major nonlinearities imposed upon the hydraulic control equipment.

In normal turbine control practice, the oil supply for the hydraulic control equipment is obtained from the same pump that supplies the lubricating oil to the turbine-generator equipment. Therefore the relatively large hydraulic equipment driving the turbine admission valves and the bypass valve mechanisms must not operate at such rates as to demand a greater hydraulic oil flow than that available from the pumps, thereby causing an undesirable pressure drop of the oil system. Since the hydraulic equipment has such physical limits upon the strokes that valves can move no further than fully closed or fully open, it is necessary to insert these limitations upon the simulation of the equipment. Another nonlinearity to include is the effect of the differential areas of the operating pistons of the valve mechanisms, as well as the effect of the static load forces imposed upon the operating pistons.

The analog computer simulation of a typical hydraulic amplifier such as is used in the pressure and speed loop representation is shown in Fig. 11. In Fig. 11(A) a much simplified schematic diagram is given of the hydraulic amplifier, showing its three basic components: the pilot valve, the operating piston, and the feedback linkage mechanism. The two expressions describing the operation of this hydraulic amplifier are given

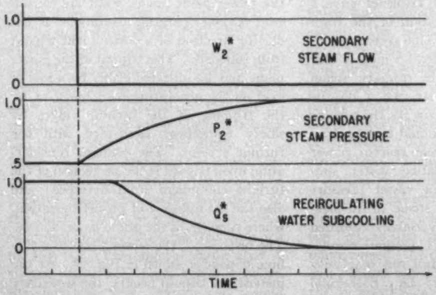
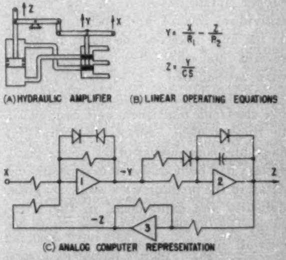


Fig. 9. Recirculating water-loop response


Z
Y
X
$Y = \frac{X}{R_1} - \frac{Z}{R_2}$
$Z = \frac{Y}{CS}$
X
-Y
Z
-Z
1
2
3

(A) HYDRAULIC AMPLIFIER (B) LINEAR OPERATING EQUATIONS

(C) ANALOG COMPUTER REPRESENTATION

Fig. 11. Hydraulic-amplifier simulation

in (B), in which the displacement of the pilot valve Y is equal to the displacement of the input X minus the displacement of the feedback from the output position Z, all multiplied by the proper scale factors which are determined by the linkage arm level ratios. The second relationship which holds is that the displacement of the output system is proportional to the integral of the disdisplacement of the pilot valve. This expression is written under the assumption that the rate of flow of hydraulic fluid to the pilot valve is directly proportional to this displacement in either direction of operation.

These two linear operating equations can be simulated by the use of three operational amplifiers as shown in Fig. 11(C). The simulation of the first equation is performed on amplifier 1, the second equation is simulated on amplifier 2, and amplifier 3 is used strictly as a sign changer. The output position of the hydraulic amplifier is obtained as the voltage output of amplifier 2, which has a silicon diode limiter connected across it in order to simulate the stroke limit of the operating piston. The input to this integrator representing the operating piston is so obtained through two adding resistors that for one polarity of input signal, one resistor alone is in the circuit, whereas for the opposite polarity input the two resistors are in parallel. This simulates the effect of the static load upon the operating piston. The output of amplifier 1, which is the position of the pilot valve, is limited in either direction of operation by the two silicon diodes connected across the amplifier. The breakdown voltages of these diodes are selected to obtain the proper rate limits upon the operating piston.

TOTAL POWER PLANT SYSTEM SIMULATION

The over-all simulation of the Dresden nuclear power plant was obtained by the proper interconnection of the various individual control and operating loops. In the final phases of the study, in which several abnormal operating conditions were investigated, up to 70 operational amplifiers, three function generators, and five electronic multiplers were committed to this problem. Since the individual loops and components of the power plant system had been analyzed in considerable detail in the preliminary studies, no great difficulty was experienced in the interconnection of the various loops into the over-all system simulation.

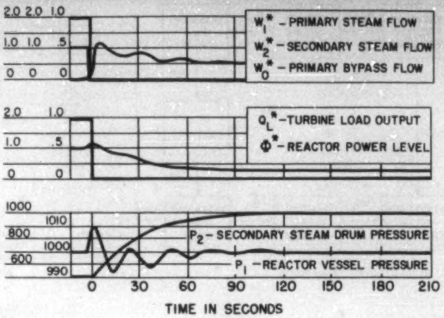


Fig. 12. Transient performance following load tripout

In the scaling of the problem for the computer, a nondimensional form of scaling was used. The electronic multipliers used in this study had scale factors such that 25 volts times 25 volts as inputs gave 25 volts output with a saturation level of approximately 50 volts in the output. Therefore the problem was scaled so that 25 volts represented 100% rated value of the different system variables in those portions of the problem in which electronic multipliers were required. In other portions of the problem 50 volts was chosen for rated conditions. This adoption of a normalized or a per-unit scaling made interpretation of the data quite easy, in that no unusual scale factors were involved. Also, the adoption of this universal scale factor of either 25 or 50 volts for rated conditions eliminated much of the risk of data being recorded at the wrong scale factor.

The immediate results of this study were obtained in the form of recorder charts of the transient performance of different system variables. Throughout most of the study, ten channels of information were recorded. This was obtained by using an 8-channel recorder with a Triplexer connected to the one channel. Since the problem was scaled for real time, the chart speed was sufficiently slow that the Triplexer gave a reasonably good indication of the transient conditions. The nine system variables which were recorded for most of the transient runs were the primary steam flow W_1, the secondary steam flow W_2, the primary bypass steam flow W_0, the turbine load demand Q_D, the actual turbine load obtained Q_L, the nuclear reactor power level Φ, the recirculating water subcooling Q_s, the reactor vessel pressure P_1, and the secondary steam drum pressure P_2. Additional variables recorded were the turbine speed ω, pressure regulator reference P_{1R}, control rod reactivity ΔK_c, accident reactivity ΔK_A, individual bypass valve steam flow W_{0R}, feedwater flow W_s, and other system variables. Several hundred transient responses were taken of the power plant performance for a large variety of driving functions. To maintain control of the accuracy of the data being obtained, a series of log sheets was set up in which each of the transient runs was given an identifying number, with the log sheet reporting such information as the driving function, the variations in the different system parameters being considered, the recorder calibration for all of the channels of information being recorded, and the pertinent information being obtained from the recorder chart. From this relatively large amount of data taken, the observed data were presented as families of curves of system performance indications such as transient pressure overshoot or time to reach certain magnitudes as functions of the various driving functions and system parameter combinations.

Indicated Transient Performance

POWER PLANT TRIPOUT FROM FULL LOAD

One of the most demanding driving functions which can be imposed upon the the power plant, under what are considered normal operating conditions, is the driving function of a power plant tripout from full load. This tripout of the power plant may be initiated either by a loss of load on the generating equipment, or by the tripping of the turbine valves by safety equipment associated with the turbine system. The primary consideration upon tripout is the protection of the turbine equipment from overspeed and also the maintaining of proper operation, where possible, of the prime power source. The response of the power plant to a turbine tripout is shown in Fig. 12. At the instant the tripout occurs, the secondary

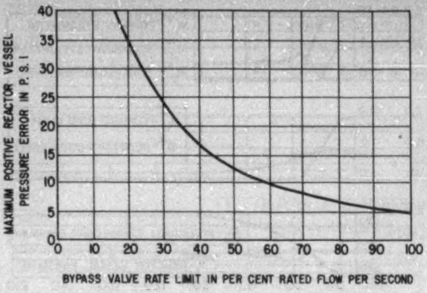


Fig. 13. Effect of bypass valve rate upon pressure transients

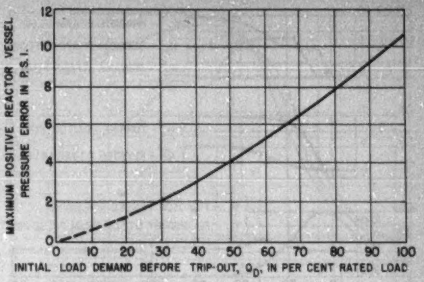


Fig. 14. Effect of tripout level upon reactor-vessel pressure transient

steam flow and primary turbine steam flow decrease to zero very rapidly. With the time constants of typical control system, these valves shut off in much less than one second's time. The primary bypass steam flow then increases up to a magnitude to maintain pressure within the reactor vessel. The rate of increase of this steam flow is determined by the rate limit upon the opening of the bypass valve and is so set that there is no detrimental pressure drop in the oil supply system. The turbine load output drops very rapidly to zero following the decrease in the two steam flows. The reactor power level takes a momentary positive overshoot and then decreases to approximately 20% power in the steady-state conditions which represent the stand-by boiling of the reactor with the control rods at the normal conditions. The primary steam drum pressure overshoots approximately 10 psi (pounds per square inch) and then goes through a damped oscillation until the nominal 1,000-psi pressure is obtained. The secondary steam drum pressure starts out at the 500-psi value of full-load conditions and then increases to the 1,000 psi, which gives the same pressure in the secondary steam drum as in the primary system.

The magnitude of the first oscillation of the pressure loop is determined primarily by the magnitude of the steam flow error signal imposed upon it. This steam error signal is measured by the rate at which the bypass valve opens following the tripout of the primary steam valve. From a series of transients in which this rate limited was varied, the curve of positive transient pressure overshoot of the reactor vessel pressure as a function of bypass opening rate is plotted as shown in Fig. 13. This figure shows that a bypass valve rate limit below approximately 20% flow change per second causes an undesirable positive reactor vessel pressure overshoot of over 35 psi. If the valve is speeded up to the magnitude that the limit is effectively removed and its speed is only limited by its inherent time constant, the positive reactor vessel pressure overshoot is held to a value less than 5 psi.

With the bypass valve rate limit set at approximately 60% flow change per second, and the initial power level as a variable, the curve of maximum positive pressure overshoot as a function of the initial load demand preceding the tripout is obtained as shown in Fig. 14. This shows that the pressure error is practically a linear function of the initial power level before the tripout has occurred.

NORMAL POWER CHANGE MANEUVERS

Two recorder chart representations are now presented which illustrate the normal maneuvers of the power plant in changing power level at a reasonable rate. In the first case, illustrated in Fig. 15, the turbine load demand is increased linearly from zero at the steady-state standby condition to 100% in 25 seconds. At this standby condition, the primary steam valve and secondary steam valve are closed; the bypass valve is allowing approximately 40% rated primary steam flow directly to the condenser; the turbine load output is at zero; the reactor power level is at approximately 20% rated power; the reactor vessel pressure is at the rated 1,000 psi; and the secondary steam drum pressure is also at 1,000 psi.

As the load demand is applied to the turbine control system, the primary steam valve opens to admit primary steam to the turbine simultaneously as the bypass valve closes. At the time that the primary steam valve is carrying all of the the primary steam flow, the secondary steam valve begins to open. The secondary steam valve opens at a rate determined by the power demand, and in this particular case reaches its full open limit of approximately 130%. When the secondary steam valve reaches its limit, there is still insufficient primary steam flow to maintain the desired load upon the turbine. Therefore it can be seen that, at the point where the secondary steam valve hits its limit, the turbine load output deviates from the demand of the 25-second ramp input and does not reach the desired 100% load until approximately 50 seconds after the demand had been initiated. The secondary steam drum pressure does not start to drop until the secondary steam flow is started. Since the recirculating water subcooling does not begin to increase immediately after the secondary steam drum pressure begins to drop because of the transport time delay, the reactor power level does not begin to increase until some time after the initiation of the driving function. This can also be seen from the curve of primary steam flow which remains at practically a constant value until nearly 25 seconds after the transient was started. It then begins to increase at a rate determined primarily by the secondary steam drum pressure time constant acting through the recirculating water subcooling to give a reactor power level value. The reactor vessel pressure is found to be only slightly disturbed by this transient, having a maximum positive deviation from the nominal value of about 3 psi and a negative deviation of less than 1 psi.

In the second case, illustrated in Fig. 16, the power plant demand is assumed to be changed from 100% to 0% load linearly over a 25-second period. The reduction of turbine load demand causes a decrease in the secondary steam flow. When the

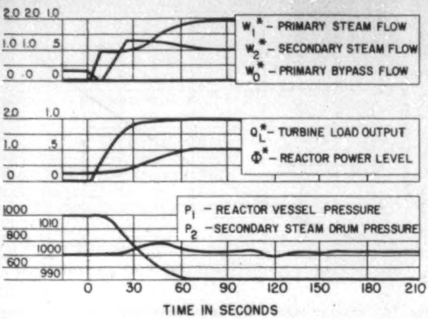


Fig. 15. Transient performance during load increase from no load to full load

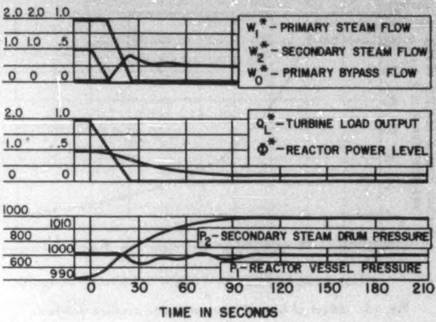


Fig. 16. Transient performance during load decrease from full load to no load

secondary steam valve has been completely closed, the primary steam valve is closed simultaneously with the opening of the bypass valve. In this case, the change in the secondary steam drum pressure begins very shortly after the initiation of the transient, since the secondary steam flow is the first of the two steam flows to change. The recirculating water subcooling begins to decrease, accompanied by a decrease in the reactor power level. It should be noticed that the turbine load output follows exactly the turbine load demand in going from 100% to 0% in 25 seconds of time. The reactor vessel pressure shows a negative error of approximately 4 psi with practically no positive pressure error at any time during the transient. After steady-state conditions are reached, the reactor operates at slightly above 20% rated power with a bypass steam flow of approximately 40% rated high-pressure steam flow.

INDIVIDUAL BYPASS VALVE CHECK-OUT

In the operation of the bypass steam system, the magnitudes of steam flow to be handled are such that one valve is not feasible; rather, a number of valves are operated in sequence. It is estimated that eight valves will be used, each of which will be capable of allowing 15% rated high-pressure steam flow when in the full open position. As one of the operating procedures of the power plant, the operation of each of these individual bypass valves must be checked periodically. Since this will apply a transient to the pressure loop of the system, an indication of the best method of checking of these individual bypass valves is desirable. Each bypass valve is opened at a certain rate, held open for a given period of time, and then closed at one-half the rate at which it was opened. The results of a typical bypass valve flow check is shown in Fig. 17. Under these conditions, the turbine load output remains relatively constant and is not shown on this reproduction of the recorder chart. In this particular case, the power plant is operating at 100% power with rated primary steam flow, rated secondary steam flow, and with no flow through the bypass valve system. The bypass valve is opened in 5 seconds, held open for $2^1/_2$ seconds, and then closed in 10 seconds. The opening of the bypass valve causes a reduction in the reactor vessel pressure which reaches a maximum negative pressure swing of approximately 8 psi. The bypass flow obtained from this check, which first appears as shown in the curve of the bypass flow, causes the pressure regulator to reduce the primary steam flow and thus maintain the reactor vessel pressure. When the primary steam flow drops, the secondary steam flow is increased by the speed governor to maintain load upon the turbine. There is an overshoot of the reactor vessel pressure at approximately this time of the transient, reaching a magnitude of approximately 10 psi over the nominal value. This causes the pressure regulator to open the bypass valve to bring the pressure back to nominal value. There are several cycles of oscillation of the primary steam drum pressure and the three steam flow valves until steady-state conditions are reached. It should also be noticed that there is a slight oscillation in the reactor power level

Since it was desirable to find the best method of checking out the bypass valves with minimum disturbance on the system, a considerable number of runs were taken in which both the rate of opening and the rate of closing the bypass valve were varied, as well as the length of time at which the valve was held open. Data were taken for three values of hold-open time of 0 seconds, $2^1/_2$ seconds, and 5 seconds and for rates of valve opening ranging from 1/2 second for full opening time down to a slow speed of 50 seconds for full opening time. The results of this study are plotted as shown in Fig. 18. In those cases in which the valve is opened quite rapidly, the primary driving function is the area under the curve, which is a measure of the total volume of steam withdrawn. It can be seen that the hold-open time is then quite important in determining the magnitude of the positive pressure error. As the rate of opening of the bypass valve is slowed down, the pressure loop is found able to follow the change in steam flow and to maintain the reactor vessel pressure. As the time of opening approaches 50 seconds, the pressure error is somewhat less than 2 psi. From all considerations, it appears that an opening time of approximately 5 seconds falls in the region of the largest transient disturbance upon the pressure loop of the system.

Conclusions

The basic theory of operation of the dual-cycle boiling-reactor power plant has been pointed out, along with examples of methods of its simulation upon the analog computer equipment, and several specific examples of transient performances as indicated from the simulation. The primary results of the study indicate the general trends of performance to be expected for the system under the assumptions being made in the analysis.

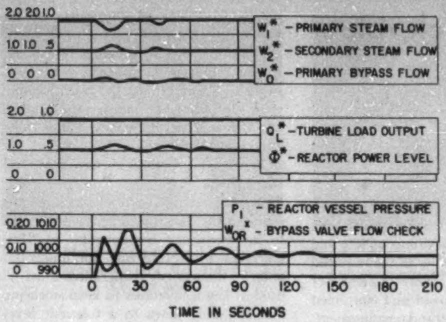


Fig. 17. Transient performance during bypass-valve check

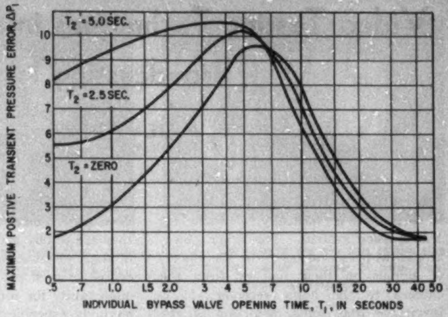


Fig. 18. Effect of bypass-valve checking procedure upon reactor vessel pressure transients

This study is not the final phase of the dynamic analysis of the particular nuclear power plant under consideration but, rather, is a step in the formulation of design characteristics and the understanding of the basic problems involved in its performance. This preliminary study points out several areas in which there is some question as to the assumptions made during the analysis of the system. In particular, two areas are being investigated in further detail for inclusion in a more comprehensive simulation of the system. These areas include a more refined thermal model of the reactor vessel and primary steam drum system in which the effects of the hydrodynamic coupling between the two vessels are included. This system includes the steam storage volumes in the dome of the reactor vessel and in the primary steam drum. The two storage volumes are coupled by a fairly long riser containing a mixture of steam and water which has the peculiar characteristic of being a relatively dense medium and at the same time being quite compressible. The effect of this coupled mass-spring system must be considered in the determination of the total power plant operation.

Another area in which additional information is needed is in the improvement of the reactor-void fraction map within the reactor core. In addition, the representation may become quite complex and include the hydraulic effects of the changing of flow velocities within the core channels as a function of the accumulation of the steam void density. Moreover, it may be necessary to go to a nodal model of the reactor in which both time and space dimensions of neutron flux are considered, rather than the calculated lumped space-time relationship with an assumed time function appended to compensate for the assumptions made in the original simulation.

The results of this study show the versatility of the analog computer technique in handling some of the extremely complex nonlinearities and product relationships which exist in large complex control systems. Although it would be feasible to program this problem for a digital type of analysis, much more data appears obtainable in a more reasonable time by using the analog computer techniques. The results are presented in such a form that a good appreciation of the physical picture of the real time performance of the system is obtained. In addition, the analog computer simulation may be adapted to form the heart of a power plant simulator, in which the outputs are used to drive the various meters and instrumentation equipment. Such a power plant simulator may be used as an operator training device so that experience may be gained in the operating characteristics of the reactor power plant long before the actual components are put in operation.

References

1. Dual Cycle Improves Boiling-Water Reactors, S. A. Untermyer. *Nucleonics*, New York, N. Y., vol. 13, no. 7, July 1955, pp. 34–35.

2. Dresden Boiling-Water Reactor, J. R. Wolcott, V. A. Elliott, G. M. Roy, E. P. Peabody, G. Sege. *Mechanical Engineering*, New York, N. Y., vol. 79, Mar. 1957, pp. 238–39.

3. Transient and Steady State Characteristics of a Boiling Reactor, The Borax Experiments, J. R. Dietrich, D. C. Layman. *AECD-3840*, Argonne National Laboratory, Lemont, Ill., 1953.

4. Boiling Reactors; Direct Steam Generation for Power, S. A. Untermyer. *Nucleonics*, vol. 12, no. 7, July 1954, pp. 43–47.

5. The HRE Simulator, An Analog Computer for Solving the Kinetic Equations of a Homogenous Reactor, D. M. Collier, L. A. Meers, J. P. Palmer. *ORNL-1572*, Oak Ridge National Laboratory, Oak Ridge, Tenn., Sept. 24, 1954.

6. Nuclear Reactor and Power Plant Simulator. *GER-7648*, Goodyear Aircraft Corporation, Akron, Ohio, May 23, 1956.

7. Analysis of the Transient Conduction of Heat in Long Solid Cylindrical Fuel Elements for Nuclear Reactors, F. E. Tippets. *HW-41895*, Hanford Atomic Products Operation, Richland, Wash., May 1, 1956.

8. Representation of Nonlinear Functions of Two Input Variables on Analog Equipment, D. A. Elliott. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, Apr. 1957, pp. 489–95.

9. Transport Delay Simulation Circuits, G. S. Stubbs, C. H. Single. *WAPD-T-38*, Westinghouse Electric Corporation, Pittsburgh, Pa., 1954.

10. Control of Nuclear Reactors and Power Plants (book), M. A. Schultz. McGraw-Hill Book Company, Inc. New York, N. Y., 1955.

Video Testing Techniques in Television Broadcasting

A. STE-MARIE
NONMEMBER AIEE

A TELEVISION video transmission system is basically a waveform transmission system. The prime objective is the preservation of the waveforms of signals on their journey from the camera to the home receiver. It is inevitable that some distortion of the signal occur in any practical transmission path. The amount of distortion which is permissible is based not on physical measurements alone but on viewing tests by critical observers. Those physically measurable characteristics in which large variations cause no detectable difference in the quality of the received picture may be permitted large tolerances, while those to which the picture quality is particularly sensitive are permitted only small tolerances.

The following transmission parameters are usually taken into consideration:

1. Amplitude versus frequency characteristic
2. Phase delay versus frequency characteristic
3. Transient response
4. Differential gain and phase

Although a substantial amount of effort has been devoted to determining the effects of these transmission factors on picture quality, especially in color transmission, a great deal more work is required before the true significance of each parameter is known as well as establishing adequate correlation between them. More complete performance specifications are urgently needed for monochrome television systems and components.

The faithful reproduction of color pictures imposes yet more stringent requirements on video transmission systems. At the outset, the National Television Systems Committee recommended standards and tolerances dealing with several of the transmission parameters. These recommendations were eventually approved by the FCC (Federal Communications Commission). Certain of the tolerances specified in the color standards, although justified from the standpoint of ultimate picture quality, are difficult to attain in practice. Tolerances for other parameters are not specified.

New video testing techniques have been developed to meet the challenge although they are still by and large incomplete and in many instances lacking in the required degree of accuracy. There is plenty of opportunity for improved and simplified methods of specification and measurement. Test procedures should be standardized in order to insure greater convenience along with completeness, repeatability, and easy communication of the test results.

About 2 years ago the tolerances in the FCC color television standards were tentatively divided as follows: 60% for the transmitter equipment, 20% for the studio equipment, and 20% for the studio to transmitter link. At that time, no reference was made to the intercity network facilities. Several industry committees, including those of IRE (Institute Radio Engineers), the Radio-Electronics-Television Manufacturers Association, and the Canadian Radio Technical Planning Board, have been investigating this problem further, but progress has been slow to date. These same committees are also active in analyzing test equipment requirements and testing procedures.

Amplitude Versus Frequency

The amplitude versus frequency response of a video transmission system is a measure of the ability of the system to reproduce detail in the horizontal direction. It can be shown, assuming an ideal phase characteristic, that the maximum steepness of transitions in the picture and therefore picture sharpness is a function of the integral of the amplitude versus frequency response of the system, i.e., the area under the curve. Picture quality is not appreciably affected by variations of the shape of the amplitude versus frequency characteristic in the vicinity of the cut-off frequency, so long as the average rate of change in this frequency range is not altered. However, in the substantially flat part of the amplitude characteristic, comparatively small variations with frequency cause echo effects in the picture.

In color television, tests have indicated that critical observers would rate 1 or 2 db (decibels) amplitude variations in the region of the color subcarrier, with respect to low-frequency components, as "Impairment to picture but not objectionable." Even smaller tolerances must be used at low frequencies to keep smearing and streaking down to a tolerable level for both monochrome and color transmissions.

It is generally advisable to use a reduced input voltage for frequency response measurements, in order to prevent overload conditions or other nonlinearities in the equipment under test from interfering with the measurement. Special care must be taken when checking systems and/or components exhibiting peaks in the frequency response.

SINE-WAVE OSCILLATORS

Point-to-point sine-wave testing is by far the simplest and possibly the most precise method of measurement. It possesses the added advantage of providing for a complete coverage of the video frequency band. However, it is a time-consuming process and one that cannot readily be used to check the performance of systems containing keyed clampers. The point-to-point test method is recommended for applications where a high order of accuracy is desirable and when time permits.

MULTIBURST GENERATORS

The multifrequency burst generator, better known as the multiburst generator, produces a series of equal amplitude bursts of sine wave of various frequencies, usually extending from 0.5 mc (megacycles) to 6.0 mc. In addition, the output wave form includes sync and blanking pulses and a white-bar signal; Fig. 1.

For quick checks, discrete sine-wave bursts of known frequencies are usually sufficient to show at a glance the general nature of the amplitude versus-frequency characteristic. Test results are easy to tabulate and report. The test signal includes sync and blanking pulses and therefore will pass in the normal manner through the various television components and circuits. The multiburst test signal is especially well-suited for

---

Paper 57-666, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 18, 1957; made available for printing July 9, 1957.

A. STE-MARIE is with the Canadian Broadcasting Corporation, Montreal, Que., Canada.

The author wishes to thank J. E. Hayes and J. Carlisle for their encouragement and valuable assistance, and Mr. Fournier for his contribution in assembling the test data and preparing the slides for this paper.

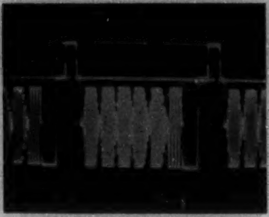


Fig. 1. Multiburst test signal. The sine-wave test frequencies are: 0.5, 1.5, 2.0, 3.0, 3.6, 4.2 mc

routine checks on systems. For this reason it has been recommended for television network tests by the Joint Television Network Committees, grouping the Canadian Broadcasting Corporation, the Trans Canada Telephone System and the Canadian National and Canadian Pacific Telegraphs; see Figs. 2–5.

Sweep Generators

Sweep generators provide a continuous display of the amplitude versus frequency response of the circuit under test. Sync and blanking pulses can be readily added to the swept sine-wave output of the generators. Combined with fixed-frequency markers, this test signal is as convenient as the multiburst test signal from the standpoint of tabulating, reporting, and comparing results. Swept signals are especially well suited for equipment and system alignment including equalizer adjustments, over-all studio system tests as well as routine checks on all video circuits; see Figs. 6–16.

Phase Delay

In television transmission systems, the times of transmission of all frequencies in the video spectrum should be substantially equal. The transmission time over a system is known as the phase delay or envelope delay and is expressed mathematically as the first derivative of the phase shift with respect to the frequency. If the phase is measured in radians and the frequency in radians per second, the phase delay will be in seconds.

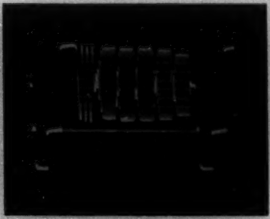


Fig. 2. Typical test result using the multiburst test signal. Circuit under test: local video channel

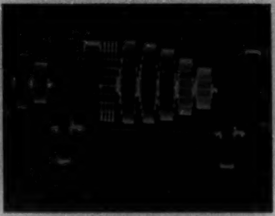


Fig. 3. Another test result exhibiting a 6-db drop in the amplitude response at 4.2 mc

The time of transmission of a single-frequency wave through a system is called the absolute phase delay or envelope delay of the system at that frequency, but it is not usually possible or necessary to measure this. The quantity that is of importance is the difference in phase delay or envelope delay between signal components of different frequencies.

In discussing phase-delay distortion in television transmission systems, we must differentiate between the parts of the frequency spectrum in which the distortion occurs. In the televised picture of a thin vertical line, for example, phase-delay distortion at the higher frequencies would be revealed by a displacement of the edges of the line and by one or more displaced images commonly referred to as ringing. But with a wider object, only its edges would be affected by phase-delay distortion at the higher frequencies.

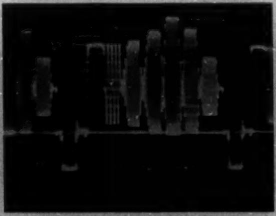


Fig. 4. Test signal as viewed at the output of an improperly aligned stabilizing amplifier. The response is overpeaked in the vicinity of 3.0 mc and falls off rapidly thereafter. There is also evidence of black compression interfering with the measurement

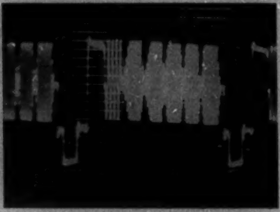


Fig. 5. Example of the a-c axis of the sine waves being shifted because of poor low-frequency response

Constant phase delay in a transmission system gives the best transient response possible for a given amplitude response. Delay distortion at the higher frequencies will also be accompanied by a corresponding decrease in the rate of rise of the transient response with a consequent loss in picture sharpness due to blurring of sharp edges. A phase-delay distortion of 0.1 μsec (microseconds) of the higher frequency components in a television signal is readily apparent, and in actual practice, the phase-delay distortion at the higher frequencies should not be allowed to exceed approximately ±0.5 μsec. Phase-delay distortions at the middle frequencies from about 200 kc upwards, affect the definition of a picture markedly. This type of distortion is better known as smearing; the impairment is a blurring of the vertical edges of objects in the televised scene and the whole picture looks as if it had been smeared along the horizontal axis. Finally, phase-delay distortions at the lower frequencies, from about 15 kc to approximately 200 kc, result in the most objectionable type of picture impairment, known as streaking.

It is quite certain that small delay distortions in the low and middle frequency band affect picture quality markedly,

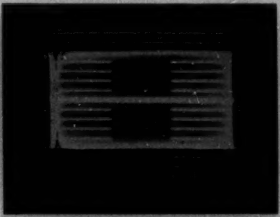


Fig. 6. Swap generator output waveform. (The single variable frequency marker is barely visible as a small pip located near the top right-hand corner of the display)

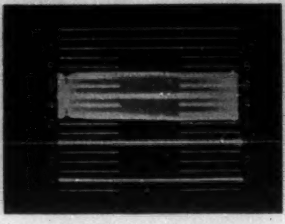


Fig. 7. Same test signal as in Fig. 6 but with added sync and blanking pulses

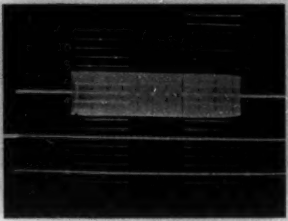


Fig. 9. Same test signal as in Fig. 8 but with added sync and blanking pulses

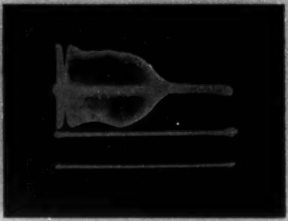


Fig. 11. Typical system test result. Circuit under test: over-all studio facility plus local video channel. Marker set at 4.2 mc

whereas they become of less importance as the upper frequency limit of the system is approached. Consequently, narrow tolerances should be placed on delay distortions in the low- and middle-frequency range.

In color television the subject of phase-delay distortion assumes additional importance because 1. misregistration of the luminance and chrominance signals in the reproduced color picture is an evident defect, 2. in receiving the color transmission on a monochrome receiver the compatibility will be impaired if the pattern resulting from the chrominance signals is not coincident with the corresponding luminance signal, and 3. abrupt delay variations occurring over rather small frequency increments in the upper portion of the video band can cause shifts in phase between various color components creating hue shifts.

It is difficult to measure relative phase delay to the required degree of accuracy. Most known methods of measurements apply only to the frequency range extending from approximately 500 kv upward. Phase-delay distortions occurring in the range from 15 kc to approximately 500 kc, cannot be measured conveniently at this time. This is indeed unfortunate since these distortions will cause impairments to the picture which are most objectionable in monochrome as well as in color transmission. It is the author's opinion, no matter what method of measurement is used, in order to insure maximum usefulness, the test signal selected should be a synthetic television signal; i.e., a test signal containing sync and blanking pulses.

Transient Response

The amplitude-frequency response of a monochrome system is geneally maintained fairly constant so as to assure a reasonable phase characteristic rather than to avoid a loss in picture detail. Although more important, the phase-frequency characteristic alone is not a complete measure of the waveform transmission capabilities of a system. A television transmission system is basically a waveform transmission system, i.e., a system in which it is important to preserve the waveforms of signals. Transmission performance is better expressed by the waveform responses to suitably chosen test signals because they can more directly be related to picture quality, even though, from a mathematical point of view the steady-state (amplitude and phase) and waveform responses are equivalent in that a full knowledge of either enables the other to be calculated. Square-wave testing was introduced early in television transmission work. Among the pioneers of waveform testing in television were Puckle in 1935 and Bedford and Fredendall in 1939 and 1942. However, progress has been slow and waveform testing and specifying has not yet become common at least on this side of the Atlantic. However, this is not true of several European countries, especially England, where waveform techniques have been developed and are now well established.

SINE-SQUARED PULSES

In the recent years, sine-squared pulses have been substituted for square waves. A sine-squared pulse has the property of containing relatively large amplitude components of frequency up to a certain frequency, and only small components of almost negligible amplitude above that frequency. If the half-amplitude duration of the pulse is equal to $1/2f$, the response is down 6 db at a frequency f. Further, the frequency spectrum is zero at $2f$. For example, if f is chosen as 4 mc (the nominal television bandwidth) the corresponding half-amplitude duration is 0.125 μsec. The spectral

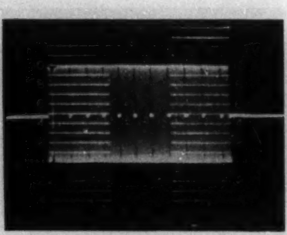


Fig. 8. Sweep generator output wave form (ten fixed frequency markers, spaced 1 mc apart)

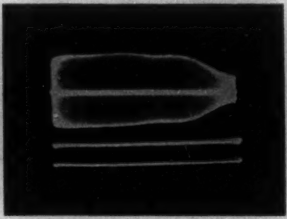


Fig. 10. Typical equipment test result. Equipment under test: camera channel amplifier. Marker set at 8 mc

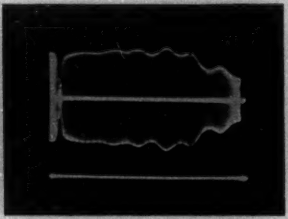


Fig. 12. Oscillations in the amplitude response caused by the mistermination of a coaxial cable run. Marker set at 7 mc

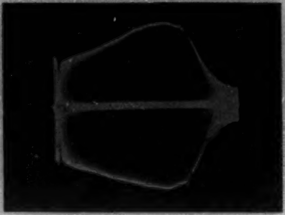


Fig. 13. Test signal as viewed at the output of an image-orthicon camera preamplifier showing the peaking introduced for purposes of aperture correction. Marker set at 7 mc

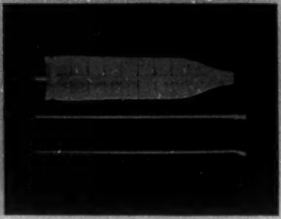


Fig. 15. Typical equipment test result. Equipment under test: stabilizing amplifier

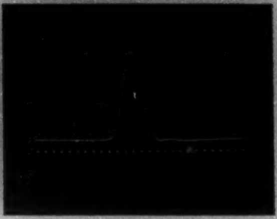


Fig. 17. Sine-squared pulse (half-amplitude duration: 0.125 μsec)

response of this pulse is 6 db down at 4 mc and is zero at 8 mc; see Fig. 17.

An ideal square wave intended for testing a system over the upper part of the frequency range generally produces a waveform response which includes a substantial amount of irrelevant information tending to obscure the relevant. The effect is minimized by using a sine-squared pulse, having an amplitude-frequency spectrum which falls away as rapidly as possible above the nominal upper cut-off frequency of the system under test. Furthermore, if the half-amplitude duration of the pulse is made $1/f$ or 0.25 μsec, its spectrum is down 6 db at 2 mc and is zero at 4 mc; see Figs. 18 and 19.

The particular merit of this latter pulse is that it is capable of being transmitted without distortion through an ideal low-pass system with an upper cut-off frequency of 4 mc. In other words, the response of a typical 4-mc television system to this pulse contains practically no irrelevant information. It is then especially well-suited for routine tests on such systems. However, these test waveforms and others can be used to advantage only if adequate correlation has been established between waveform response characteristics on the one hand and picture quality on the other. Ideally the performance specifications for any component or part of a system should be based on subjective picture appraisal determinations of the maximum amount of waveform distortion tolerable in a complete television system and subsequent allocation of fractions of this amount to eack link. In actual practice, due attention must be given to results known to be attainable with some regard to economy. Once the performance specifications are agreed upon, waveform responses can be evaluated at a glance. Suitable oscilloscope masks can be devised for this purpose. Among the features of sine-squared pulse responses usually specificed are the maximum allowable half-amplitude duration and the maximum tolerable overshoots and undershoots, leading, and trailing (Figs. 20–26).

Although wave-form techniques are remarkably simple and meaningful, there remains a serious drawback. Should the transient response fall outside the limits specified, it usually becomes necessary to revert to steady-state measurements in order to secure quantative data on the amount of amplitude and/or phase distortion present. This is so, mainly because, in the present status of the art, components and systems are specified throughout the industry in terms of steady-state responses and alignment procedures are based on this more conventional approach. However, the newly developed waveform synthetizing and waveform-correcting networks of the "time equalizer" and other types will allow one to operate directly on the waveform response. Another obstacle to a more rapid progress of waveform measuring techniques has been the difficulty of handling the computational problems of cascade connected links and networks. This difficulty has now been greatly overcome by the development of the time-series method.

Time series can also be used to advantage in the derivation of the steady-state characteristics from the transient response. Greater speed and convenience as well as a much higher order of accuracy can be achieved by this method. However, waveform responses secure direct information on the waveform transmission capabilities of the circuit under test. On the other hand, measurements of amplitude and phase versus frequency have little significance unless the results can be interpreted in terms of waveform

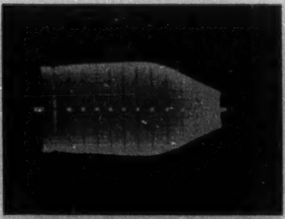


Fig. 14. Typical equipment test result. Equipment under test: master control switching facility

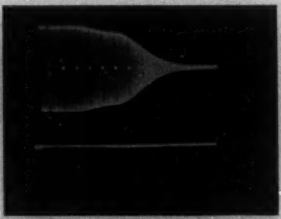


Fig. 16. Typical system test result. Circuit under test: over-all studio facility

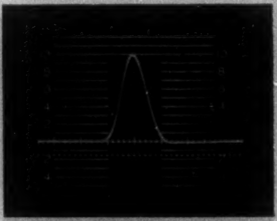


Fig. 18. Sine-squared pulse (half-amplitude duration: 0.250 μsec)

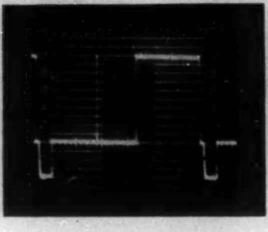


Fig. 19. Complete test signal combining the sine-squared pulse and the white bar

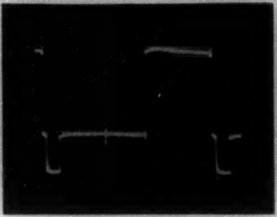


Fig. 22. Response of the same video local channel to the complete test signal, including the 4-mc pulse and the white bar

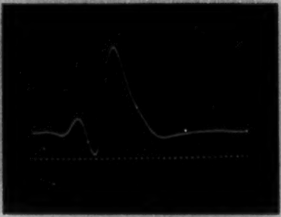


Fig. 25. Response of a long-distance television network channel to the 4-mc sine-squared pulse

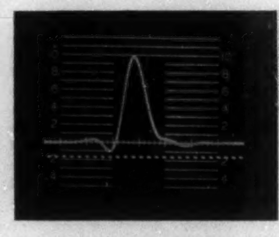


Fig. 20. Response of a video local channel to the 4-mc sine-squared pulse

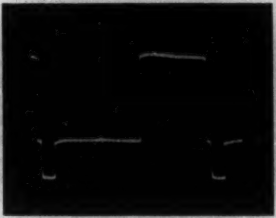


Fig. 23. Same as Fig. 22 but with 2-mc pulse

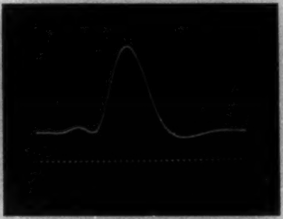


Fig. 26. Response of the same long-distance television network channel to the 2-mc pulse

responses. Therefore, for our needs, the derivation of the steady-state characteristics from the transient response is a matter of little more than purely academic interest.

WINDOW-TEST SIGNAL

Rectangular pulses, at field and line rates, are now in common use for checking the low-frequency transient response of television components and systems. A typical test signal is the "window" signal, so-called because it produces a white window on a dark background when displayed on a picture monitor; see Fig. 27. Originally developed to help locate streaking, the window-test signal will provide a ready measure of the transmission performance at the lower end of the video spectrum. It is an extremely useful tool in television transmission work especially in view of the lack of suitable methods of measurement of the amplitude and phase responses in that portion of the video spectrum. The window-test signal embodies sync and blanking pulses and can therefore be used to check the over-all as well as any part of the television system.

As was pointed out earlier, streaking is caused by transmission irregularities in the frequency region up to about 200 kc and will therefore be detected by the use of the window signal. Streaking is a typical television picture defect that affects almost equally both color and black and white signal transmission.

Distortions that cause streaking can occur in any part of the transmission system from the camera to the television receiver. Prevention of these defects required very close control of transmission characteristics in the lower frequency portion of the video band, see Figs. 28–30. A 5-degree phase shift or 1-μsec delay at the fundamental frequency will cause a tilt of approximately 24% in the window-test signal corresponding to an

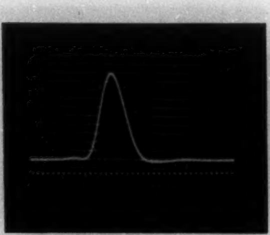


Fig. 21. Response of the same video local channel to the 2-mc pulse

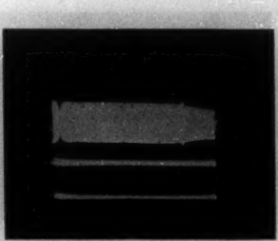


Fig. 24. Amplitude response of the same video local channel. Marker set at 4.2 mc

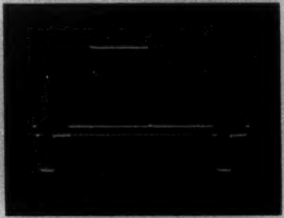


Fig. 27. Window generator output waveform

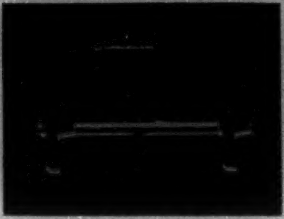


Fig. 28. Typical test result. Approximately 2% waveform distortion. Circuit under test: local video channel

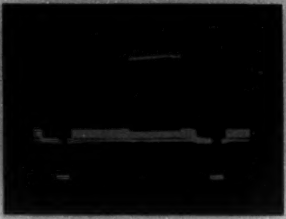


Fig. 30. Example of positive streaking. Amount of tilt: 3%

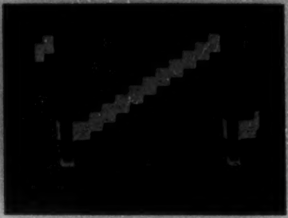


Fig. 32. Output of stairstep generator. Low-frequency exploratory signal plus high-frequency sine wave

intolerable amount of streaking in the transmitted picture. As a matter of fact, a 5% tilt will result in distortions readily noticed during normal picture transmission, and defects corresponding to as little as 2% tilt may be detectable under certain conditions. It is worth noting that while a 5-degree phase shift gives 24% tilt, the change in amplification that accompanies this in a normal resistance-capacitance coupled amplifier can be shown to be about 0.05 db.

Differential Gain and Differential Phase

The following definitions have been extracted from the IRE Standards on Television; definitions of Television Signal Measurement Terms, 1955. Differential gain: in a video transmission system, the difference in the gain of the system for a small high-frequency sine-wave signal at two stated levels of a low-frequency signal on which it is superimposed. Differential phase: in a video transmission system, the difference in phase shift through the system for a small high-frequency sine-wave signal at two stated levels of a low-frequency signal on which it is superimposed.

The term level is used here to define a specified position on an amplitude scale applied to a signal wave form. If the foregoing definitions are accepted, it is redundant to speak of differential gain distortion or differential phase distortion, since differential gain or differential phase is distortion. Both differential gain and differential phase are zero for an ideal transmission circuit. In monochrome television, differential phase has little significance if any. Differential gain, on the other hand, is a measure of the departure from linearity of the transfer characteristic of the transmission system. In severe cases of differential gain this form of distortion will result in either white compression or black compression or a combination of both. Small amounts of differential gain have but little effect on picture quality.

In color television transmission, it is an entirely different story. Since the transmitted luminance component and the chrominance components should be independent of each other, it is evident that the chrominance components must not be affected by the luminance. Otherwise, for example, as an actress walked from the dimly lighted edge of a scene to the brightly lighted center the color of her dress would change strangely. The color subcarrier signal component being transmitted by adding it to the luminance or monochrome component, it follows that the average luminance level determines the *a-c* axis of the subcarrier.

Therefore the desired independence of chrominance and luminance signals means that the change of gain of the subcarrier as a function of signal level, and likewise the change of phase, should be zero, in other words, the differential gain and the differential phase should both be zero. Differential gain will cause a change in saturation of the colors while differential phase will be responsible for a change in hue. Of the two types of distortion, the latter is obviously the more objectionable. The average viewer is usually critical of hue shifts, espcially when flesh tones and objects of readily identified colors are being reproduced. It is well known that a phase shift of about ±5 degrees at the color subcarrier frequency is readily detectable.

The definitions presented in the foregoing for differential gain and phase contain within themselves a suggestion as to a basic measurement technique. Differential gain and phase can be measured with the aid of a test signal having two essential components, a low-frequency "exploratory" signal of sufficient amplitude to explore the entire amplitude range of the system, plus a high-frequency sine-wave of relatively small amplitude superimposed on the exploratory signal. The

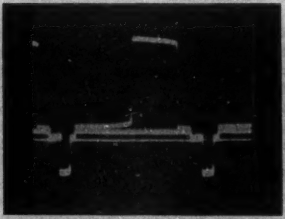


Fig. 29. Example of negative streaking. Amount of tilt: 6%

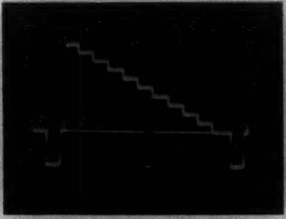


Fig. 31. Output of stairstep generator. Low-frequency exploratory signal only

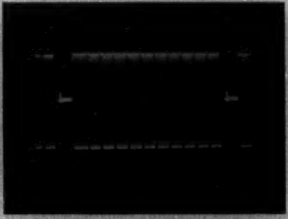


Fig. 33. Same as Fig. 32 but after having passed through a high-pass filter

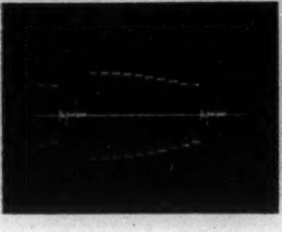


Fig. 34. Differential gain measurements. Approximately 33% differential gain (black compression)

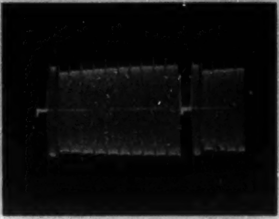


Fig. 35. Differential gain measurements. Approximately 15% differential gain (white compression)

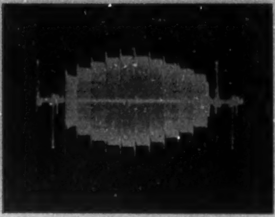


Fig. 37. Typical oscilloscopic display when both white and black compression are present

test signal should preferably be a synthetic television signal, containing synchronizing and blanking pulses, which can be passed through all normal transmission circuits; see Figs. 31 and 32. For differential gain measurements, the output of the equipment under test is applied to an oscilloscope scanning at half the line frequency and through a high-pass filter which will remove the low-frequency exploratory signal and leave the subcarrier oscillations, see Figs. 33–35.

At this time, we should like to stress an apparent deficiency in the definitions of differential gain and phase quoted previously. As will be remembered, the definitions refer to the difference in gain and phase of the system for a small high-frequency sine-wave signal at two stated levels of a low-frequency signal on which it is superimposed. Whatever these two stated levels are, say blanking and reference white levels, the measured amounts of differential gain and differential phase at these two levels may yield an incomplete and inaccurate picture of the transmission performance of the circuit under test. As a matter of fact, it can be shown that under certain conditions, relatively small amounts of differential gain and phase can be measured between two stated levels of the low-frequency signal while it is evident that large amounts of differential gain and phase actually exist when the over-all amplitude range of the picture signal is taken into consideration; see Figs. 36 and 37.

The authors would like to recommend the following alternate definitions. Differential gain: In a video transmission system, differential gain is the maximum difference in gain of the system for a small amplitude high-frequency sine-wave, the a-c axis of which is varied from blanking to peak-white level by means of a low-frequency exploratory signal on which it is superimposed. Differential phase: In a video transmission differential phase is the maximum difference in phase shift through the system for a small amplitude high-frequency sine wave the a-c axis of which is varied from blanking to peak-white level by means of a low-frequency exploratory signal on which it is superimposed.

Differential phase can be measured by means of a high-pass filter coupled to a synchronous detector, in which the phase of the subcarrier component of the test signal is compared with that of a continuous reference subcarrier signal. A cathode-ray oscilloscope can be used as a null indicator at the detector output, and phase differences between the unknown and the reference subcarrier can be measured with a calibrated phase shifter. The reference subcarrier is usually extracted from the test signal itself.

Calibration of Test Equipment

The need for adequate calibration of the test equipment cannot be too strongly emphasized. Voltmeters and oscilloscopes must be checked for voltage calibration. This may be done conveniently, using a suitable substandard voltmeter, specified to the required order of accuracy. More important still is the calibration of the amplitude versus frequency response of the oscilloscope. A useful device for the calibration of oscilloscopes is the Western Electric type *70B* transmission measuring set. It is a thermocouple operated power measuring meter designed to operate in the frequency range from direct current to 10 mc. The accuracy in terms of absolute power expected at 0 db on the meter scale is ±0.1 db, while the differential accuracy over the entire frequency range is within +0.1 db. Such an instrument is not only extremely useful for the calibration of oscilloscopes, but will permit making direct point-to-point measurements on individual amplifiers, equalizers, cables, and other system components whenever a higher degree of accuracy may be required.

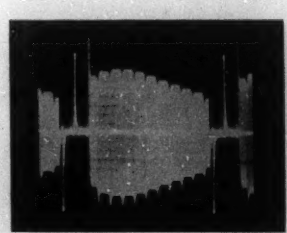


Fig. 36. Typical oscilloscopic display when both white and black compression are present

The frequency calibration of multiburst generators, sweep generators, and sine-wave oscillators should be checked with adequate care. One will understand that test results can hardly be correlated if the test frequencies cannot be ascertained within reasonable limits. Frequency calibration can usually be performed to a sufficient degree of accuracy without recourse to expensive and bulky frequency measuring equipment.

Oscilloscopes used for waveform testing especially sine-squared pulse measurements must be carefully calibrated along the time axis. Marker generators used for this purpose should preferably be crystal controlled.

In all, it is well to remember that even the best of test equipment is susceptible to changes in performance and drifts in calibration, if not the occasional failure. For this reason, the calibration of the test equipment must be checked regularly, in order to insure accurate and reliable test results.

References

1. **Routine Tests and Alignment Procedures for Amplifiers and Circuits Used for Color TV, Howard C. Gronberg.** ***Engineering Conference,*** **National Association of Radio and Television Broadcasters, Chicago, Ill., May 1954. (Available from NBC Engineering Department.)**

2. Apparatus for the Measurement of Phase Delay in Television Transmission Circuits and in Associated Equipment, C. W. Goodchild, R. C. Looser. *Proceedings*, Institution of Electrical Engineers, London, England, 1952, vol. 99, pt. IIIA, pp. 788–95.

3. Color-Video Envelope-Delay Measurement, R. C. Kennedy, H. French. *Electronics*, New York, N. Y., June 1955, pp. 144–48.

4. Color Test Equipment and Test Procedures, John W. Wentworth. *Broadcast News*, Radio Corporation of America, Camden, N. J., vol. 90, Aug. 1956, pp. 46–51.

5. Color Video Tester Checks Distortion, H. P. Kelly. *Electronics*, New York, N. Y., Sept. 1954, pp. 128–31.

6. Analysis, Synthesis and Evaluation of the Transient Response of Television Apparatus, A. V. Bedford, G. L. Fredendall. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Oct. 1942, pp. 440–57.

7. Spectrum Analysis of Transient Response Curves, H. A. Samulon. *Ibid.*, Feb. 1951, pp. 175–86.

8. A Theory of Time Series for Waveform Transmission Systems, W. E. Thomson. *Monograph no. 53*, Institution of Electrical Engineer, London, England, Oct. 15, 1952, pp. 397–409.

9. Waveform Computations by the Time-Series Method, N. W. Lewis. *Proceedings*, *Ibid.*, vol. 99, pt. III, 1952, pp. 294–306.

10. Waveform Responses of Television Links, N. W. Lewis. *Ibid.*, vol. 101, pt. III, July 1954, p. 258.

11. A Testing Pulse for Television Links, I. F. MacDiarmid. *Ibid.*, vol. 99, pt. IIIA, 1952, pp. 436–44.

Statistical Encoding for Text and Picture Communication

W. S. MICHEL
NONMEMBER AIEE

SUGGESTIONS have been made from time to time for reducing the transmission time or bandwidth required for the facsimile transmission of text or pictorial material by representing it by a description which has low redundancy. This paper presents an investigation of the channel capacity required for various types of copy, using one such description, based on one-dimensional differential coordinates together with statistically matched encoding. A binary digital transmission channel is assumed throughout. Only black and white copy is considered. A written or printed message is assumed to consist of a 2-dimensional array of resolution elements, called dots, whose number determines the definition. Since a definition of 0.01 inch is widely accepted as satisfactory for facsimile reproduction of black and white copy, this figure, corresponding to almost a million dots for an $8^1/_2$- by 11-inch typewriter-size sheet, is used throughout.

Dots, Dot Groups, and Patterns

A piece of copy may be reproduced at another location if a tabulation is available specifying for all the resolution elements or dots whether they are black or white. Such a tabulation may be called a "description" of the copy. The description need not necessarily proceed dot by dot and we shall distinguish among three schemes of description, i.e., three types of "alphabets." These are in terms of dots, dot groups, and patterns respectively. The first and most elementary specifies blackness or whiteness for each dot individually. In the second, configurations of dots, i.e., dot groups, are specified. For example, the description may take the form of a statement of the number of black, or white, dots occurring in successive continuous runs measured along one co-ordinate, with the other held constant;[1] or it may have more elaborate forms. In the third scheme, patterns of black or of white dots are the alphabet items and the description of the copy is in terms of these patterns and their locations. Patterns, as defined here, are assumed to include only a limited number of select configurations characteristic of the specific class of material for which the system is designed, but not all the possible configurations. A system based on a pattern alphabet cannot, therefore, handle copy in which patterns occur which it does not "recognize." On the other hand, the dot system, and the dot-group system as here defined, can recognize any and all copy, subject only to the limits set by the resolution.

Depending on the alphabet chosen and, of course, the subsequent coding used, the channel capacity required for average copy may be large or small. The actual information content of the copy is in general lower than this, the difference being the redundancy. For maximum efficiency of information transfer the alphabet and the coding should be such that there is a minimum of redundancy.

In practice, the efficiency that can be attained is usually limited by cost considerations. Thus the choice of an alphabet must be scaled to one's ability to build readers and printers able to operate in terms of the alphabet; and the degree of statistical matching may have to fall short of ideal in order to keep translator complexity down. Readers for recognizing printed characters can vary from fairly simple to rather elaborate computer-like designs, depending on the constraints one is willing to put on such factors as shape or position. More sophisticated systems operating with whole words or word groups would probably be more complex, unless messages are readily identifiable and small in variety, such as greetings telegrams. The ideal system which would recognize each piece of copy as a unit, and send only one code character for it, appears of forbidding complexity at this time, except for simple cases.

At the other end of the spectrum from these sophisticated approaches is the familiar commercial facsimile. Here each dot is, in effect, specified individually, and the reading and printing processes may be regarded as consisting simply of the identification and production of black dots, one at a time. Efficiency tends to be low, each piece of material taking the same transmission time per unit area. The differential co-ordinate description to be discussed uses a simple kind of dot-group alphabet which lends itself to a fairly straightforward analysis. Combined with even simple forms of statistical encoding this description makes it possible, in general, to specify an $8^1/_2$- by 11-inch page by considerably fewer than a million binary digits.

Differential Co-ordinate Encoding

Command Structures

In differential co-ordinate or "run-length" encoding, the number of black, or white, dots occurring in continuous runs of

---

Paper 57-723, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 19, 1957; made available for printing June 14, 1957.

W. S. Michel is with Bell Telephone Laboratories, Inc., New York, N. Y.

The author wishes to thank B. R. Kretzmer, W. A. Depp, and W. O. Fleckenstein, all of Bell Telephone Laboratories, for many valuable discussions.

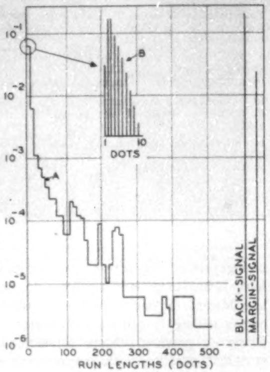


Fig. 1. Probability distribution of run lengths in a 300-word pica-typewritten letter

one color along scan lines is measured by a scanner. Such a scanner may be thought of as operating like a television or facsimile scanner modified so that whenever it encounters a transition from black to white or from white to black it sends to the encoder a count representing the number of dots between this and the previous transition. Upon receiving such a signal, the encoder translates it and subsequently sends the appropriate digital code character over the line, possibly via a storage device.

Table I. Portion of Pica-Tailored Code

Item	Code Symbol
Black	00
2 dots	010
3 dots	011
4 dots	100
5 dots	1010
6 dots	1011
Margin (also empty line)	11000
7 dots	11001
1 dot	11010
8 dots	11011

Table II. Binary Digits per 8½- by 11-Inch Page

Type of Copy	Binary Digits
Empty page	5,500
Simple drawing, Fig. 3(B)	50,000
Grid with 2-inch spacing	60,000
300-word pica letter	100,000
Circuit schematic, Fig. 3(A)	120,000
Page filled with pica type	300,000
Grid with 1-millimeter spacing	400,000
Page filled with 8-point type	700,000
Two black and one white dot alternating horizontally	3,000,000

In general, each code character represents a command to the receiver. Such commands may be specifications of run lengths or of operational routines such as "print the following length black," or "move to the next line." The ensemble of different commands pertaining to a specific scheme of description will be called a "command structure."

Various command structures are possible within the differential co-ordinate family. These may differ in such secondary features as the method of distinguishing black from white lengths, or in the way advantage is taken of the fact that black lengths always follow white, and vice versa. In the course of the present work, it was found that several such variations gave rise to differences in efficiency of the order of only 10%. Therefore, results are quoted only for one specific command structure. This has the following characteristics:

1. Black and white runs of the same magnitude have the same code characters.
2. The receiver assumes runs to be white, unless they are preceded by a special code character which indicates otherwise.
3. Single black dots occurring, for example, in the vertical legs of typed characters are not sent. Instead, the receiver is engineered to print a black dot automatically after each white run.

Typical Probability Distributions

The probability distribution on which a code is based should be that which will be encountered, on the average, in the copy to be transmitted. No account is taken here of the error statistics of the transmission medium, or of the relative seriousness of errors in the various code characters. For a general system, it appears to be more important to provide good matching for text than for drawings, since text is likely to contain finer detail, and hence requires greater channel capacity, as measured in the differential co-ordinate description. Therefore, most of the results here presented are derived for a "pica-tailored" language. This means a statistical encoding based on the probability distribution of commands as encountered in pica-typewritten material, in the form of an English-language business letter.

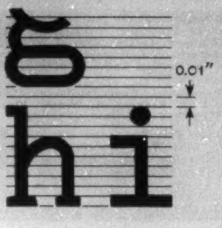


Fig. 2. Enlargement of pica-typewriter characters

This distribution exhibits a high peak at very short run lengths, and subsequently a rapid drop off, as shown in curve *A* of Fig. 1 in which the probability

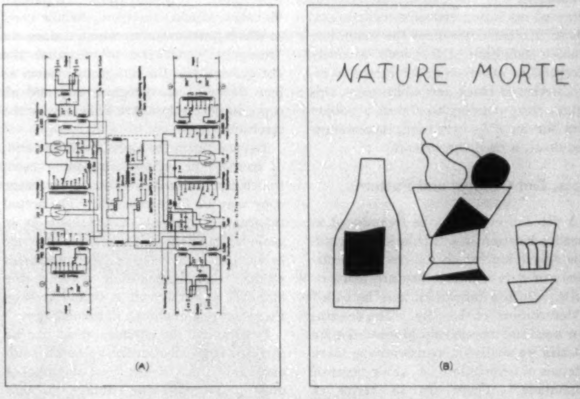


Fig. 3. Two types of drawings

A—Circuit schematic
B—Simple drawing, for which channel capacities are given in Table II.
Frames indicate size of 8½- by 11-inch sheet

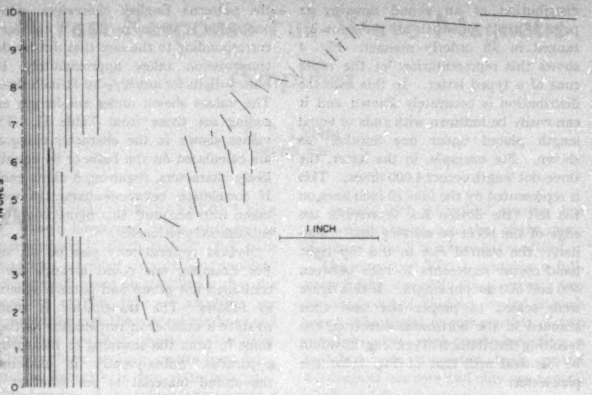


Fig. 4. White run lengths in a 300-word pica-typewritten letter

per command, averaged over 10 commands at a time, is plotted as a histogram. Curve *B* represents the probabilities of the first ten commands, plotted individually. Minor peaks occur in the probability curve at the lengths corresponding to the distances from the edge of the paper to the left margin, to paragraphs, etc. The exact nature and position of these will vary with the particular spatial arrangement of the copy; but their effect is negligible since they represent low probability totals.

Throughout this work, the scanning is assumed to proceed across the page at right angles to its longer dimension. The actual measurements of run lengths were made on enlargements of the 26 lower-case letters of a pica typewriter. For run lengths associated with single letters and for run lengths occurring between adjacent letters, the measured values were weighted in accordance with the well-known monogram and digram frequencies, respectively, of the English language.[3] For runs between alternate letters, such as between the top of the *b* and the top of the *t* in *but*, trigram tables[4] give some of the required probabilities of occurrence. The remaining values were determined by actual count through several pages of typewritten material. A typical enlargement of pica-typewritten letters is shown in Fig. 2.

For drawings or mixtures of text and drawings, the longer runs become relatively more probable. Black dots occur in a more random manner than in type so that an exponential distribution of run lengths might be expected. Other arguments lead to a hyperbolic distribution. In any event, short runs still tend to predominate over the long ones. Moreover, as mentioned above, the number of run lengths per unit area is smaller, so that run-length encoding is advantageous even with the less accurate statistical matching resulting from using the pica-tailored code for drawings.

CHOOSING A CODE

The encoding problem consists of setting up a code so that the expected length per code character is a minimum. A well-known procedure for this is the so-called Shannon-Fano code and its modification by Huffman.[3-6] The results presented here are based on a Shannon-Fano encoding of the command structure described previously using a smoothed form of the probability distribution shown in Fig. 1. Because of the large alphabet involved, a close match is possible, yielding an actual average length per code character only about 1% higher than the ideal length, i.e., the entropy. Correlation between successive run lengths is not taken into account in this encoding. The initial portion of the code book is shown in Table I. Code-character lengths are found to vary from 2 binary digits for the black signal to 20 digits for the signal corresponding to the maximum run length.

RESULTS FOR TYPEWRITTEN TEXT

With the code described in the previous section, approximately 100,000 binary digits are found to be required for the description of a 300-word pica letter. In view of the good statistical matching obtained, this also represents the lower bound on the number of binary digits required to express the average page in terms of individually coded run lengths. The same figure may also serve as an upper bound on the actual information content, which must therefore be less than 100,000 bits.

The foregoing results may be extended to other kinds of type by qualitative considerations. For example, it is readily seen that the required number of binary digits will be approximately proportional to the number of characters, and further that the effects of differences in spatial arrangement will, in general, be small.

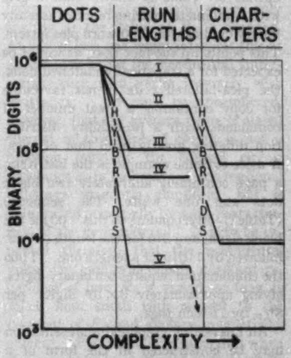


Fig. 5. Channel capacities required for various types of facsimile transmission, versus complexity of instrumentation

I—Solid page of 8-point print
II—Solid page of pica type
III—300-word pica letter
IV—Simple drawing
V—Empty page

Table III. Categories for Encoding Picture Material

	Dots	Dot Groups	Patterns
Required channel capacity	generally high: same for all types of material	variable; lower for simple or favored material	very low for special material
Versatility	unlimited	unlimited	handles special material only

Narrower characters will tend to require a similar, and wider characters a somewhat higher binary digit rate per character than pica.

RESULTS FOR DRAWINGS

The differential co-ordinate description of drawings would be expected to exhibit features differing from that for text. For example, one would expect a greater fraction of long runs, as has been pointed out, but a smaller total of runs. Different drawings will not, in general, present such high uniformity in their differential co-ordinate descriptions as different pieces of text. Therefore, it is not as easy to make general and yet fairly precise statements about binary digit rates for drawings as for text. However, individual drawings often have features which make a reasonable estimate possible by inspection or by a few quick measurements. One such feature which is often found is the occurrence of the same run length many times, as in a rectangular block.

Channel capacities required for some types of drawings are listed in Table II in terms of the total number of binary digits. These figures are based on the pica-tailored code. It may be noted that the schematic, Fig. 3(A), while much more complicated than the simple drawing Fig. 3(B), requires only about twice as many binary digits, chiefly because its many small run lengths match the code more closely.

Comparison of the fourth and seventh entries of Table II shows that a grid with 1-millimeter spacing requires as many binary digits as four 300-word pica letters. This points up the fact that, as would be expected for a statistically matched code, the pica-tailored code is not favorable for copy containing a great number of commands with a probability distribution differing greatly from that of pica. A more extreme example is the last item, a page containing alternately two black dots and one white. The sequence (Table I) corresponding to this is 00 for the black signal and 010 for the length of two, followed by 11010 for a length one. Thus the combination requires ten binary digits, giving approximately 3×10^6 digits per $8^1/_2$- by 11-inch page.

An interesting graphical representation may be constructed in the form of a drawing having the same run-length distribution as an actual drawing or page of text, but with all the runs arranged in an orderly manner. Fig. 4 shows this representation for the white runs of a typed letter. In this case the distribution is accurately known and it can easily be redrawn with runs of equal length placed under one another, as shown. For example, in the letter, the three-dot length occurs 4,000 times. This is represented by the four 10-inch lines on the left (the dotted line represents the edge of the sheet or starting line); similarly, the slanted line in the top right hand corner represents 15 runs between 400 and 500 dots in length. If this figure were scaled to proper size and then scanned in the horizontal direction, the resulting distribution of run lengths would be identical with that of Fig. 1 for the pica letter.

Discussion

While this study suggests that run-length encoding is capable of yielding substantial reduction in transmission time as compared with dot-by-dot transmission, a number of facts should be pointed out. First, in run-length encoding, an error in one binary digit tends to propagate and hence is more serious than in the dot-by-dot description where it merely changes the polarity of one dot. Second, the present study is limited to a very simple form of differential co-ordinate encoding. Correlation between successive runs is not taken into account and the encoding is one dimensional, no advantage being taken of the very pronounced regularities in the vertical direction exhibited by many types of copy.

In Table III the various approaches to the encoding of picture material which have been discussed or touched upon are reviewed. As one proceeds from left to right, required channel capacity decreases, but instrumentation complexity increases. Versatility is unlimited in the first two categories in that all possible combinations of black and white dots are permitted; in the pattern category it is limited to the allowed patterns. Fig. 5 shows semiquantitatively the different channel capacities required for certain types of material for the specific case where the dot groups are run lengths and the patterns English characters. Only one value is shown in the dot category, corresponding to the fact that dot-by-dot transmission takes approximately 10^6 binary digits for any $8^1/_2$- by 11-inch sheet. The values shown under run-length encoding are those from Table II. The values shown in the character category are calculated on the basis of 64 equally likely characters, requiring 6 digits each. If correlation between characters were taken into account[6] this figure could be substantially reduced.

Hybrid systems may also be set up. For example, one could arrange for a combined dot group and pattern scheme as follows. The transmitter is made to store a number of run lengths as they come in from the scanning or measuring apparatus. Subsequently it examines the stored material to see whether it makes up one of a number of predetermined patterns. If so, a special code symbol for the pattern is sent; if not, the code symbols for the stored run lengths are sent routinely. An alternative example, a hybrid scheme falling between the dot and length categories would be provided by applying to facsimile a coding procedure discussed by Shannon.[6] This is suitable for a source with an alphabet of two items of very unequal probabilities and might be applied to facsimile by making black and white dots the two items. While run-length encoding is only a first and simple step in this spectrum of possibilities, the results obtained are believed to be encouraging.

References

1. CODING PROCESSES FOR BANDWIDTH REDUCTION IN PICTURE TRANSMISSION, A. E. Laemmel *Report R 246–51*, Microwave Institute, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., Aug. 1951.

2. SECRET AND URGENT (book), Fletcher Pratt. Blue Ribbon Books, New York, N. Y., 1939.

3. A MATHEMATICAL THEORY OF COMMUNICATION, C. E. Shannon. *Bell System Technical Journal*, New York, N. Y., vol. 27, July 1948, pp. 379–473; Oct. 1948, pp. 623–56.

4. THE TRANSMISSION OF INFORMATION, R. M. Fano. *Technical Report No. 65*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass., 1949.

5. A METHOD FOR THE CONSTRUCTION OF MINIMUM-REDUNDANCY CODES, D. A. Huffman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Sept. 1952, pp. 1098–1101.

6. PREDICTION AND ENTROPY OF PRINTED ENGLISH, C. E. Shannon. *Bell System Technical Journal*, New York, N. Y., vol. 30, Jan., 1951, pp. 50–64.

A Fast-Response Full-Wave Magnetic Amplifier

G. E. LYNN
MEMBER AIEE

A FEW RECENT PAPERS[1-3] have been concerned with analyses of conventional, single-ended, full-wave magnetic amplifier circuits (or "doublet" amplifiers[4]) based on the theory that the reactors of this type of amplifier are coupled through the control windings and that this coupling results in feedback. The feedback, in turn, causes time delay and an increase in the volts/volt gain. The purpose of this paper is to present a novel circuit which clearly demonstrates the existence of the feedback phenomenon and to offer, for the magnetic amplifier design engineer, a mathematical treatment of this phenomenon both in the time domain and the frequency domain.

Circuit Operation

If feedback does occur in doublet amplifiers, it should be possible, by decoupling the reactors during appropriate periods, to produce a circuit identical with a doublet amplifier except that feedback is absent. To this end, the circuit of Fig. 1 was devised. Use of the bridge circuit avoids the coupling of the reactors through the load circuit which occurs in the doubler and center-tapped circuits. The circuit presented in the second figure of reference 1 achieves the same end result; i.e., a fast-response full-wave amplifier, but does so by feeding back a volt-second area which is invariant with respect to the signal. Such a method introduces some complications which the circuit of Fig. 1 avoids.

To achieve circuit simplicity, the magnetic amplifier of Fig. 1 obtains its gating voltage from a square-wave generator. This eliminates the need to cascade clipping circuits to derive, from a sinusoidal source, a synchronous square wave with which to drive the switched-mode[5] transistors used. Conclusions drawn from the use of this circuit are applicable to circuits with sinusoidal sources, but some wave shapes, e.g., that of the resetting current, will be somewhat modified.

The circuit of Fig. 1 contains a full-wave bridge magnetic amplifier with two identical control windings wound on each core. The signal circuit consists of two parallel paths, each path including a switched-mode transistor, both control windings of one core and only one control winding of the other core.

In discussing circuit operation, the following simplifying assumptions will be made:

1. For the core material used, the ratio of remanent flux density to maximum flux density is unity, $B_r/B_m=1$.
2. The rectifying elements in the load circuit are ideal; i.e., zero conduction in the reverse direction and zero voltage drop in the forward direction.
3. The resistance in the windings of the reactors is negligible.
4. When applied to winding N_{G1} or N_{G2}, the magnitude of E_{ac} is just sufficient to drive the flux of the respective reactor from negative to positive saturation in one half-cycle of supply frequency.
5. With respect to a half-period of the supply, the switching time of the transistors is negligible; i.e., V_1 and V_2 open or close instantaneously as the polarity of the square-wave generator changes.
6. When operating as a closed switch, the voltage drop across the transistor is zero; when operating as an open switch, conduction through the transistor is zero.
7. The time during which "back-firing" of the rectifiers occurs is negligible with respect to a half-period of the supply.
8. The output impedance of energy sources is negligible.

The square-wave generator supplies voltage to transformer T_1 and to the load circuit of the magnetic amplifier. The center-tapped secondary of T_1 supplies base-drive current to switched-mode transistors V_1 and V_2 during alternate half-cycles, closing V_1 while opening V_2 and closing V_2 while opening V_1. When the polarity of the generator output is as shown in Fig. 1, reactor L_1 is in its gating half-cycle. Transistor V_1 is acting as a closed switch and the signal voltage, E_s, is applied to the series combination of N_{F1}, N_{S2}, N_{S1}, and R_S. In addition, during this time, transistor V_2 is acting as an open switch and the path containing winding N_{F2} is therefore an open circuit. Under these conditions, the voltages induced in windings N_{F1} and N_{S1} by the application of gating voltage to winding N_{G1} cancel each other as far as their effect on the control circuit is concerned. Because of this cancellation, the effective control circuit consists only of E_s, N_{S2}, and R_S throughout the time that the output of the square-wave generator has the polarity shown. When the polarity changes, reactor L_2 starts to gate, transistor V_1 is switched to its open condition, transistor V_2 is switched to its closed condition, the voltages appearing across windings N_{S1} and N_{F2} are in series opposition, and the effective control circuit consists only of E_S, N_{S1}, and R_S. It is now obvious that, during the reset half-cycle of each reactor, there is only one energy source to produce reset. That source is E_S. As is well known, the average volt-second area appearing across the terminals of any one winding of a reactor is zero in steady state; however, by switching windings N_{F1} and N_{F2} in and out of the control circuit at the proper intervals, each reactor presents a d-c load to the signal source.

STATIC GAIN

In Fig. 2, data are presented which show the dependence of load voltage, E_L, on E_S for various values of R_S. It can be seen from Fig. 2 that the slope of the transfer characteristic (representing the incremental volts/volt gain) varies only slightly as the value of R_S is changed. At first glance one might interpret this as an indication that the incremental volts/-ampere-turn gain varies inversely with resistance. It must be remembered, however, that the average value of signal current in this type of circuit is not a function only of the control circuit resistance as in half-wave or conventional full-wave magnetic amplifiers, but is determined principally by some characteristic of the core material. It is to be expected therefore that the incremental volts/volt gain will remain constant regardless of how R_S is varied, so long as R_S is kept small with respect to the quotient of E_S and average magnetizing current. Fig. 2 also illustrates that when the restriction on allowed values of R_S is less valid, a change in incremental volts/volt gain is discernible. For instance, changing R_S by a factor of two from 250 to 500 ohms decreases the slope by about 1.2%, whereas changing R_S by the same factor from 2,500 to 5,000 ohms decreases the slope about 11%. This indicates an effective average resistance

---

Paper 57-773, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24-28, 1957. Manuscript submitted March 22, 1957; made available for printing July 9, 1957.

G. E. Lynn is with Westinghouse Electric Corporation, Baltimore, Md.

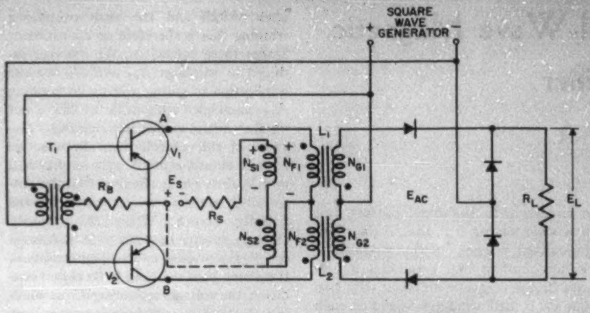


Fig. 1. Circuit of fast-response full-wave magnetic amplifier

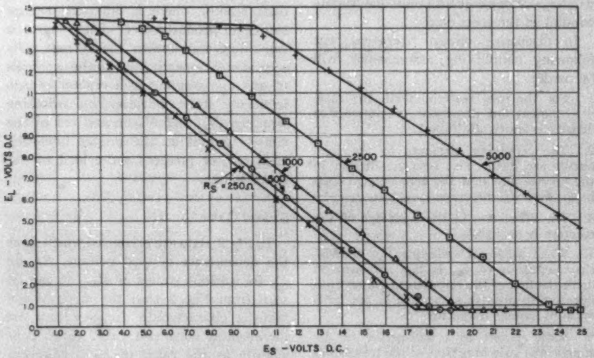


Fig. 2. Static transfer characteristics of fast-response amplifier (R_s as parameter)

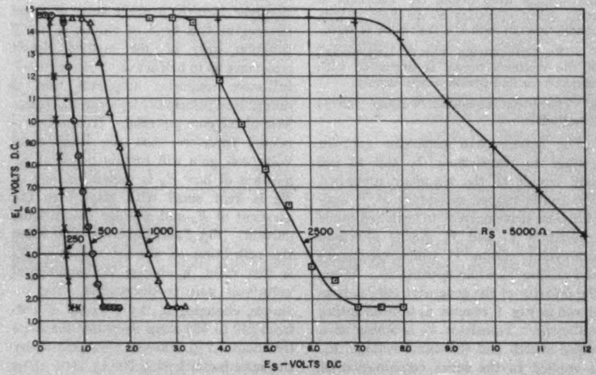


Fig. 3. Static transfer characteristics of doublet amplifier (R_s as parameter)

of between 25,000 and 30,000 ohms for the reactors.

Of additional interest in Fig. 2 is the variation, as R_S is changed, in the magnitude of signal voltage which must be applied before an appreciable amount of reset is obtained. This variation clearly indicates the existence of threshold ampere-turns[4] or a knee on the effective flux-current loop. The existence of this knee indicates that some minimum current must flow to initiate reset but, after establishing this exciting current, the amplifier is essentially a voltage-controlled device (for low values of R_S). The fact that the magnitude of this knee is proportional to the value of R_S only for larger values of R_S is attributed to the presence of some resistance other than that ascribed to R_S in the experimental control circuit, probably in the signal source used.

For purposes of comparison, the volts/volt characteristics of a conventional full-wave magnetic amplifier are shown in Fig. 3 for the same values of R_S as were used in obtaining the data of Fig. 2. The changes to the circuit of Fig. 1 to obtain a doublet amplifier are simply opening points *A* and *B* and connecting a jumper as shown by the dotted line. The effect of the presence in the control circuit of the voltage across the control winding caused by gating of each reactor is clearly evident from a comparison of Figs. 2 and 3 with proper regard to the different abscissa scales used. The incremental volts/volt gain for the same R_S is greater for the doublet amplifier in all cases. In addition, the incremental volts/volt gain of the doublet amplifier is roughly inversely proportional to the value of R_S as would be expected from consideration of the amplifier as an ampere-turn controlled device.

A significant aspect of a comparison between results obtained with the circuit operating as a fast-response amplifier and then as a doublet amplifier is that the average voltage observed across R_S is the same for both configurations, at least as long as the value of R_S remains low. A few points taken with R_S equal to 500 ohms in the fast-response configuration are shown as heavy dots in Fig. 3. The abscissa would read "average volts across R_S" for these points, the ordinate remaining the same. These data show that the volts/ampere-turn gain remains the same for both circuit types.

Dynamic Response of the Doublet Amplifier

To establish the existence of a phenomenon which can properly be called "feedback," it is necessary to show that,

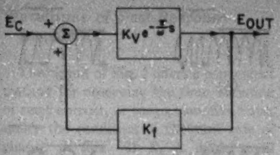


Fig. 4. Block diagram representing the 2-winding doublet amplifier

not only is the static gain affected, but the time response as well. The data of Figs. 2 and 3 have proved that coupling between the two reactors in the control circuit increases the volts/volt gain. Experimental evidence showing the fast-response nature of the circuit of Fig. 1 will be presented later and it is well known that the conventional bridge circuit does not exhibit this characteristic. However, a mathematical treatment showing why, with feedback, the response to a step input changes from a pure "dead-time" to a more or less exponential change in output will be of interest.

Johannessen has published[3] the block diagram of Fig. 4 and, using difference equations, has derived, from this diagram, an approximate volts/volt transfer function in terms of the turn ratio and various resistances. To obtain this result, an operating flux-current loop with vertical sides was postulated. While the method may be extensible to conditions where the slope of the sides of the loop is noninfinite, the mathematics would probably become unwieldy. For this reason, the treatment presented here will be expressed in terms of circuit gains only, without regard to how these gains are obtained. The treatment is rigorous and the only assumption made about the operating flux current loop is that B_r/B_m is unity. This assumption makes it possible to avoid considering voltages induced in the control circuit of the gating reactor when it is supposedly saturated. A further assumption will be made about the circuit operating conditions. It will be assumed that the circuit is biased to the upper knee of the left flank of the operating flux-current loop. This assumption in no way detracts from rigor but makes it possible to consider the output change proportional to the input without introducing another delta in the notation.

In Fig. 4, the forward path contains a block which represents the basic half-wave amplifier having a static volts/volt gain, K_v, and a dead-time, π/ω. In servomechanism notation, this block can be expressed mathematically as $K_v e^{-\pi s/\omega}$. The feedback path contains a volts/volt gain (which is less than unity), K_f. The voltage fed back is summed with the signal voltage, E_c, as shown. It must always be borne in mind that with no input, E_{out} will be a maximum. However, the voltage fed back through the gating core is proportional to the change in output voltage, ΔE_{out}, and not to the output voltage magnitude.

In the treatment, the notation below, following Johannessen as closely as possible, will be used:

E_c = control voltage averaged over a half-cycle of supply
E_{out} = output voltage averaged over a half-cycle of supply
e = base of Naperian logarithms
K = closed-loop voltage gain
K_f = feedback voltage gain
K_v = forward voltage gain
s = complex frequency variable
ω = frequency of supply, radians per second

Ignoring the frequency dependent terms for the moment, the following may be written directly from Fig. 4.

$$\Delta E_{out} = K_v(E_c + K_f \Delta E_{out}) \tag{1}$$

from which can be derived

$$\Delta E_{out}/E_c = K_v/(1 - K_v K_f) = K \tag{2}$$

which is the closed-loop voltage gain.

An objection might be raised at this point that, since a half-wave amplifier resets on one half-cycle and has an output on the succeeding half-cycle, no output voltage is available to be fed back during the reset half-cycle. Fig. 4, however, represents a doublet amplifier in which the resetting core is represented by the forward transmission and the gating core by the feedback path. Each core then alternately serves in each capacity and the gating core provides the output voltage in the proper time period.

If we assume a step input of E_c at a time when reactor L_1 is starting its reset half-cycle, no change in output voltage will appear for one half-cycle of the supply. Since both cores have previously been at saturation, no voltage appears on the windings of reactor L_2 during this half-cycle. In the next half-cycle, reactor L_1 will gate and the average value of the output voltage during this half-cycle will be reduced by an amount proportional to the voltage, E_c. Observe that the behavior of reactor L_1 during the first full cycle is identical with what it would have been if the reactor were in a half-wave circuit.

We may write:

$$\Delta E_{out(1)} = K_v E_c \tag{3}$$

where the numerical subscript indicates the half-cycle in which the parameter is measured. During this first gating half-cycle, the change in output voltage, which is equivalent to the amount of voltage which was absorbed by the reactor prior

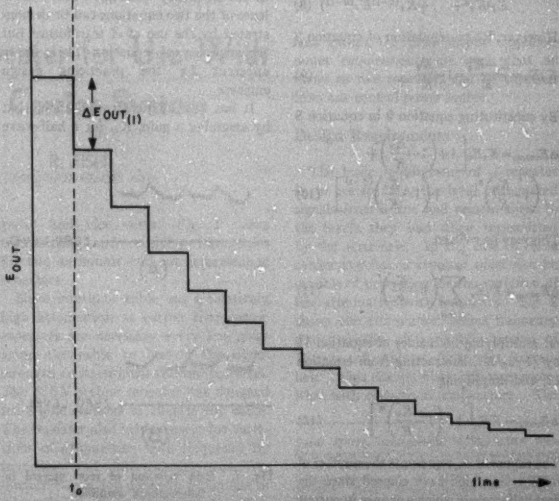


Fig. 5. Time domain response of doublet amplifier to step input applied one half-period before t_0

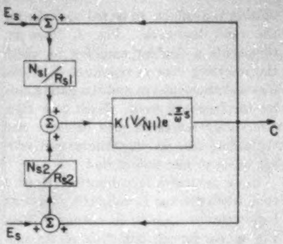


Fig. 6. Suggested block diagram representation of multi-input doublet amplifier

to firing, is reflected into the control circuit and adds to E_c in producing reset of reactor L_2. If K_f is the gain of the transfer of volt seconds from one reactor to the other through the control circuit

$$\Delta E_{out(2)} = K_g(E_c + K_f \Delta E_{out(1)}) \quad (4)$$

Substituting equation 3 in 4

$$\Delta E_{out(2)} = K_g(E_c + K_f K_g E_c) = K_g E_c(1 + K_f K_g) \quad (5)$$

Similarly

$$\Delta E_{out(3)} = K_g(E_c + K_f \Delta E_{out(2)}) \quad (6)$$

$$= K_g E_c(1 + K_f K_g + K_f^2 K_g^2) \quad (7)$$

whence, by extension

$$\Delta E_{out(n)} = K_g E_c\,(1 + K_f K_g + K_f^2 K_g^2 + \ldots + K_g^{(n-1)} K_f^{(n-1)}) \quad (8)$$

However, by manipulation of equation 2

$$K_f K_g = \frac{K - K_g}{K} = 1 - \frac{K_g}{K} \quad (9)$$

By substituting equation 9 in equation 8

$$\Delta E_{out(n)} = K_g E_c \left[1 + \left(1 - \frac{K_g}{K}\right) + \left(1 - \frac{K_g}{K}\right)^2 + \ldots + \left(1 - \frac{K_g}{K}\right)^{(n-1)}\right] \quad (10)$$

This may be written

$$\Delta E_{out(n)} = K_g E_c \sum_{r=0}^{n-1} \left(1 - \frac{K_g}{K}\right)^r \quad (11)$$

or, multiplying both sides of equation 11 by $(1-K_g/K)$, subtracting from equation 11, and simplifying

$$\Delta E_{out(n)} = K E_c \left[1 - \left(1 - \frac{K_g}{K}\right)^n\right] \quad (12)$$

The exponent, n, represents the number of half-cycles which have elapsed since the first half-cycle in which a change in output occurred following the step change in E_c. If π/ω is the time duration of a half-cycle, $n = t\omega/\pi$ where t is time measured from the beginning of the first half-cycle. Substituting for n, equation 12 can be expressed as a function of time and further simplified to

$$\Delta E_{out} = K E_c(1 - e^{-at}) \quad (13)$$

where

$$a = -\omega \ln(1 - K_g/K)/\pi$$

or

$$\Delta E_{out}/E_c = K(1 - e^{-at}) \quad (14)$$

Equation 13 is recognized as the familiar exponential usually used to represent the time delay of magnetic amplifiers. Since $\Delta E_{out(n)}$ is defined as an average quantity (averaged over π/ω), equation 13 has no meaning except at points where $t\omega/\pi$ is equal to an integer. The proper representation, then, is a series of steps, constantly decreasing in step magnitude as shown in Fig. 5.

By inclusion of the frequency dependent terms in equation 2, the following is obtained

$$\frac{\Delta E_{out}}{E_c} = \frac{K_g e^{-s\pi/\omega}}{1 - K_f K_g e^{-s\pi/\omega}} \quad (15)$$

This expression in the frequency domain is equivalent to equation 14 in the time domain (with the appropriate restrictions on allowable values of $-at$ in the time domain, and if a step input is postulated in the frequency domain). The equivalence of the two equations can be demonstrated by the use of Z transforms[6] but the meaning of equation 14 is more apparent for the practicing design engineer.

It has been demonstrated, then, that by assuming a gain, K_g, for a half-wave amplifier, a higher gain, K, is obtained for the doublet amplifier, but at the expense of an exponential form of time delay in addition to the "dead-time" inherent in all self-saturating magnetic amplifiers.

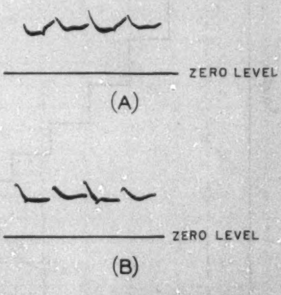


Fig. 7. Time variation of reset current in fast-response amplifier

A—One-third full output
B—Two-thirds full output

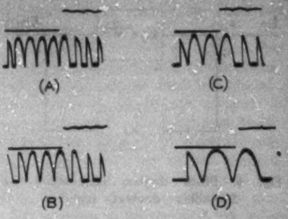


Fig. 8. Response of fast-response amplifier to square-wave input

A—80-cps input
B—100-cps input
C—133.3-cps input
D—200-cps input

THEORY LIMITATIONS

The preceding work has been concerned with an amplifier which has only two windings, a control winding, and a gating winding. The positive feedback voltage from the gating reactor adds directly to the signal voltage in the control circuit and the feedback voltage gain is a function of the control circuit resistance. However, magnetic amplifiers are frequently used as summing amplifiers because of the isolation which can be achieved by use of multiple input windings. The representation of the amplifier as a volts/volt gain device in these instances seems somewhat awkward since a reactor cannot sum, but discriminates, voltages.[7] For this reason, a representation similar to that in Fig. 6 would seem preferable. In this diagram, the basic half-wave amplifier is assigned a gain in volts/ampere-turn and the individual winding turns and control circuit resistances are indicated in separate blocks. Unfortunately, in contrast to a 2-winding reactor, the volts/ampere-turn gain of the basic amplifier block will no longer be constant as N^2/R is varied, but will be a function of the resistance values chosen and the square of the turns of the individual windings. In view of this limitation, Fig. 6 does not seem to be a satisfactory final solution and further investigation of block diagrams suitable for use in these applications is indicated before detailed analysis with this type of representation can be attempted.

Waveforms of Fast-Response Amplifier

The circuit of Fig. 1 offers a convenient method for observing the time variation of reset current, or magnetomotive force, under the condition of constant rate of change of flux. If R_s is made small (25 ohms or less) the voltage across R_s is of the order of 1/10% of the applied voltage and may be said to represent the current required to reset a reactor from a voltage source. An oscilloscopic presentation of this voltage is shown in Fig. 7. A recent paper[8] presents the core-reset function under just such conditions obtained by other methods but does not include a picture of the current for comparison with Fig. 7.

It can be seen in Fig. 7 that there is a surprisingly large difference between the shapes of the currents required to reset the two cores. The two cores were chosen on the basis of identical readings obtained from a constant current flux-reset type core tester. On the basis of average output values, the cores seem to be well-matched even under the conditions imposed by the circuit of Fig. 1, but the time variations of the currents in achieving the same amount of reset are hardly equivalent. Neither core exhibits the constant current characteristic which would be required under the assumption of a vertical-sided flux-current loop.

To illustrate the fast response nature of the circuit of Fig. 1, Fig. 8 is presented. The circuit used in obtaining these pictures differed in some respects from the circuit presented in this paper, principally in that a sinusoidal source was used in place of the square-wave generator and a square-wave generator was used to supply E_s. It can be seen that full response to the step change in signal is always achieved by, at latest, the second gating half-cycle following the step change, not counting the half-cycle in which the change is initiated. This is the best response obtainable with a magnetic amplifier since the reactor which is gating at the time of the signal change will respond completely to the change in its next gating half-cycle, but the reactor which is resetting at the time of the signal will only respond completely in its second gating half-cycle following the change.

Conclusion

A full-wave fast-response magnetic amplifier circuit has been presented which can be used to substantiate the theory that a two-winding doublet amplifier can be represented by the block diagram of Fig. 4. The limitation of Fig. 4 to 2-winding amplifier representation has been indicated and a possible path for future investigation for a more general diagram has been suggested. The restriction of signal current to the current flowing through a reactor resetting under the desired N^2/R conditions suggests that the circuit can be used as a flexible investigational tool to explore core properties.

References

1. FAST RESPONSE WITH MAGNETIC AMPLIFIERS, D. G. Scorgie. *AIEE Transactions*, vol. 72, pt. I, 1953, pp. 741–49.

2. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

3. A THEORETICAL AND EXPERIMENTAL STUDY OF MAGNETIC AMPLIFIERS, Thorleif Knutrud, Jr. *B.A. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., June 1954.

4. A NEW APPROACH TO SELF-SATURATING MAGNETIC AMPLIFIERS, T. J. Pula, F. G. Timmel. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956.

5. JUNCTION TRANSISTORS USED AS SWITCHES, R. L. Bright. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 111–21.

6. THE ANALYSIS OF SAMPLED-DATA SYSTEMS, J. R. Ragazzini, L. A. Zadeh. *Ibid.*, vol. 71, pt. II, 1952, pp. 225–32.

7. A SIGNAL-DISCRIMINATING MAGNETIC AMPLIFIER, G. E. Lynn, J. F. Ringelman. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 97–102.

8. CORE-RESET FUNCTIONS IN MAGNETIC-AMPLIFIER ANALYSIS, PART I—A CORE-RESET FUNCTION, G. C. Feth. *Ibid.*, vol. 77, pt. I, 1958 (*Paper no. 57-95*).

A Transistorized Repeater for Use With the 45BN Cable Carrier System

V. BABIN
ASSOCIATE MEMBER AIEE

R. FISH
ASSOCIATE MEMBER AIEE

THE REPEATER described in this paper is a part of the *45BN* carrier telephone system. This system transmits and receives 24 voice channels over two pairs in a nonloaded exchange or toll cable. It is a member of a family of miniaturized carrier systems using many interchangeable subassemblies.[1,2,3] The complete family of type *45* carrier equipment provides carrier transmission over open-wire cable and radio circuits.

Type *45BN* equipment uses single-side-band suppressed carrier modulation and operates in the frequency range from 40 to 140 kc (low group) in one direction and from 164 to 264 kc (high group) in the other direction. A terminal that transmits the low group receives the high group and vice versa. Fig. 1 shows transmitting and receiving frequencies for two terminals without intermediate repeaters.

Since multipair cable has a relatively high attenuation at carrier frequencies, repeaters are necessary every few miles along the cable to restore the signal strength to its original transmitted level. The *45BN* system repeater was designed for 8-mile sections of 19-gage toll cable. The repeater also helps correct for variations of attenuation with frequency by inverting and interchanging the frequency bands for opposite directions of transmission from the repeater (frequency-frogging).

By the use of transistors for all amplifiers except the final output amplifier, power requirements are such that as many as five repeaters can be supplied from one central power source.

Design Requirements

The basic requirement of a repeater is to accept the low-level attenuated signals from a line and restore them to the levels they had when transmitted by the terminals. In the case of cable carrier systems, a repeater must also be capable of correcting for the variation of line attenuation with frequency. Fig. 2 shows the attenuation versus frequency characteristics of an 8-mile section of 19-gage toll cable at 46 degrees Fahrenheit. This characteristic will vary somewhat with changes in temperature. The

---

Paper 57-907, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 28, 1957; made available for printing July 29, 1957.

V. BABIN and R. FISH are with the Lenkurt Electric Company, San Carlos, Calif.

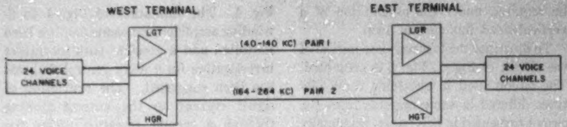


Fig. 1. Block diagram showing directions and frequency bands used for two 45BN terminals without repeaters

LGT = low group transmitting
HGT = high group transmitting
LGR = low group receiving
HGR = high group receiving

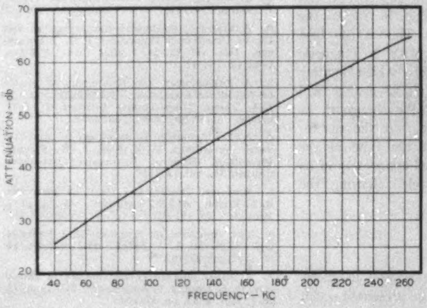


Fig. 2. Frequency versus attenuation curve for 8 miles of 19-gage toll cable. Temperature is 46 degrees Fahrenheit. Attenuation is higher at higher temperatures and lower at lower temperatures

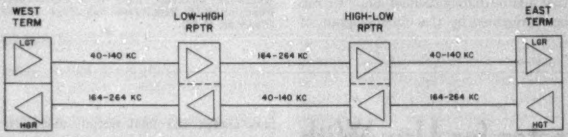


Fig. 3. Block diagram showing interchange of frequency bands by frequency-frogging repeaters

basic design requirement of the *45BN* repeater was to compensate for the effect of the attenuation characteristic on the high- and low-frequency bands transmitted by the system.

Since the attenuation characteristics vary with the temperature, regulation must be provided to adjust the gain of the repeater to maintain a constant output signal level with varying input signal level. The regulator action must be slow enough to give any system failure alarm time to operate in case of total loss of signal. Otherwise, the gain would increase too fast and amplification of noise and crosstalk would prevent the alarm from operating.

Because of the relatively short distances between repeaters, an economic advantage could be realized by feeding power remotely to repeaters. Power supplies then would not be needed at every repeater location. This would require low power consumption for each repeater, in itself, an economic advantage.

Since outdoor cabinets would, in many instances, be used to house the repeaters, the repeaters should be as small as possible to minimize the size or number of cabinets required. In addition, the repeaters must be capable of operating in ambient temperatures ranging from −40 to +60 C (degrees centigrade).

In practice, not every repeater section can be exactly 8 miles of 19-gage cable. Therefore, provisions must be made for the repeater to "see" the electrical characteristics of this design section even when the actual section differs in length or type of cable.

The *45BN* system is designed to operate in the same cable with Western Electric types *N* and *ON* systems. In some cases Western Electric *N* repeaters have been used with *BN* systems. To maintain the flexibility of application for types *45BN*, *N*, and *ON*, it was desirable that the *BN* repeater be electrically and physically interchangeable with the type *N* repeater.

Design Approach

As shown in Fig. 2, the repeater would require a gain of 64 db (decibels) in the high-group direction (the transmission loss at 264 kc) and a gain of 45 db in the low-group direction (the transmission loss at 140 kc). Equalization must correct the 19-db slope in the low group (26 db at 40 kc to 45 db at 140 kc) and the 15-db slope in the high group (49 db at 164 kc to 64 db at 264 kc). The high gain and equalization requirements could be eased considerably by using a frequency-frogging repeater.

A frequency-frogging repeater transposes the frequency bands for both directions of transmission. As the bands for both directions of transmission pass through the repeater, they are translated and interchanged to the portion of the frequency spectrum occupied by the other. The channels within the bands are also inverted in their order in the frequency spectrum. A frequency-frogging repeater can be a low-high repeater or a high-low repeater. That is, it can either receive the low groups from both directions and transmit high groups, or it can receive the high groups from both directions and transmit the low groups. In practice, low-high and high-low repeaters alternate along the cable between terminals. A block diagram of a *45BN* system with two such repeaters is shown in Fig. 3.

As a result of the interchange and inversion of channels, channels that suffer the greatest attenuation in the high group, suffer the least attenuation in the low group. For example, the channel transmitted in the 260- to 264-kc slot in the high group falls in the 40- to 44-kc slot in the low group. Therefore, the gain at each repeater need only equal the average loss between 40 and 264 kc, or 46 db. This is almost the requirement for low-group transmission only.

Frequency-frogging also makes the system almost completely self-equalizing as shown in Fig. 4. The residual slope amounts to about 2 db per repeater section. This is far easier to correct than the 19-db slope that occurs in the low group or the 15-db slope that occurs in the high group.

In addition to reducing the amplification required for individual repeaters,

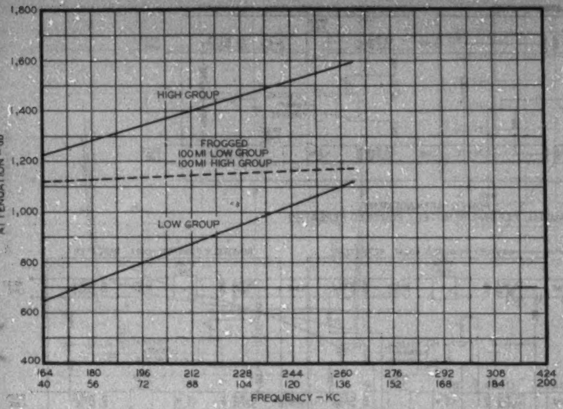


Fig. 4. Losses in 200 miles of 19-gage cable showing equalizing effect of frequency-frogging

frequency-frogging reduces run-around crosstalk. This is the crosstalk which arises through the high-level output of one repeater being coupled to a tertiary pair. The crosstalk signals are transmitted along the tertiary pair and then coupled to the low-level input of the repeater or to the inputs of adjacent repeaters. Since the frequency band transmitted by a frequency-frogging repeater differs from the frequency band received, these crosstalk signals will be out-of-band to the directional filters at the inputs.

Considerable advantage in maintenance and spare parts requirements could be realized if one repeater unit could be used interchangeably for either high-low or low-high operation. Therefore, the general design aimed at a circuit which could be changed from one method of operation to the other by no more than a minor adjustment.

The small size and low power requirements of the *45BN* repeater pointed to the use of transistors. Therefore, the development approach was based on a completely transistorized circuit. Earlier experience had shown that transistors could serve many of the functions of vacuum tubes with longer life expectancy and less power consumption.[3] It seemed natural to apply those advantages to repeaters which might be located at remote sites where minimum attendance and low power consumptions are desirable.

Laboratory investigation showed that transistors could be used in all the required circuits with one exception. This exception was the output stage of the power amplifier. The amplifier not only had to be a broad-band amplifier covering the frequency range from 40 to 264 kc, but also had to have power-handling capabilities similar to the Western Electric *408A* vacuum tube. It was at first assumed that the cost of a transistor amplifier of this type would be within economic limits. However, no single transistor was found that would consistently fulfill these requirements.

The final design of the repeater includes two vacuum tubes: one for the output stage for each direction of transmission. But the use of transistors in all other circuits of the repeater allows remote power feed to as many as two repeaters in tandem on each side of a locally powered repeater. This means that a power source is required only at every fifth repeater location.

Operation

Fig. 5 is a block diagram of the *45BN* repeater. The repeater may be connected for either low-high or high-low operation by means of threaded coaxial connectors at the front of the unit. When arranged for low-high operation, channels from the west in the 40- to 140-kc band enter the repeater through the input transformer and receiving span pad. This band then passes through the slope adjust network where 4 db of negative slope is introduced. It then passes through the 140-kc low-pass filter where unwanted high-group crosstalk (from the outputs of this and other repeaters) is removed.

The 40- to 140-kc band modulates the output of a 304-kc 2-stage transistor oscillator. The 304-kc carrier and the 40- to 140-kc input band are balanced out in the modulator. The output of the modulator then contains the two principal bands of 164 to 264 kc and 344 to 444 kc. A band-pass filter selects the 164- to 264-kc band and rejects the other.

By the time it reaches the output of the band-pass filter, the signal has been attenuated approximately 14 db: 4 db in the slope equalizer and 10 db in the modulator and filters. The 24-channel band is then amplified 60 db (46 db above the received level) and transmitted to the line.

The amplifier consists of three sections: (1) a 19-db fixed-gain single-stage transistor amplifier; (2) a 1-db, ±8 db, single-stage transistor regulating amplifier; and (3) a 40-db fixed-gain output amplifier consisting of a transistor pre-

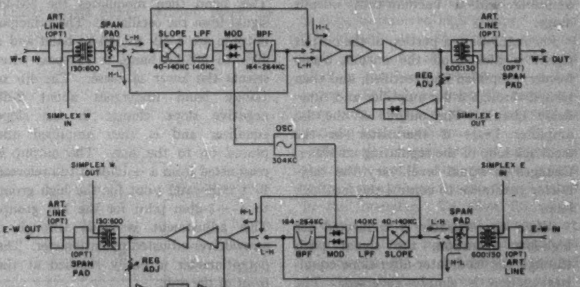
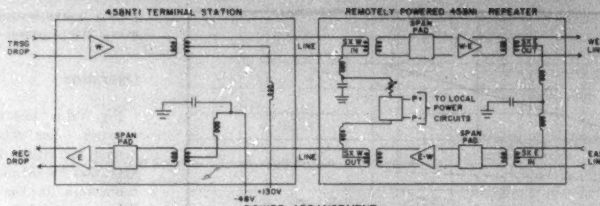
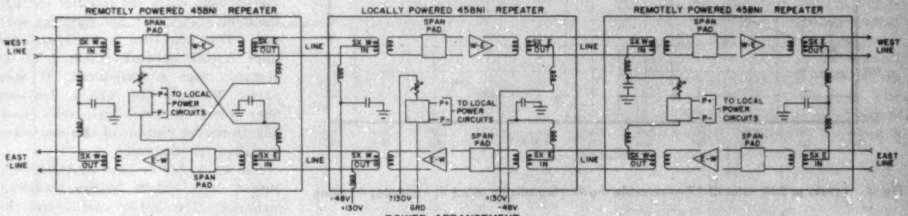


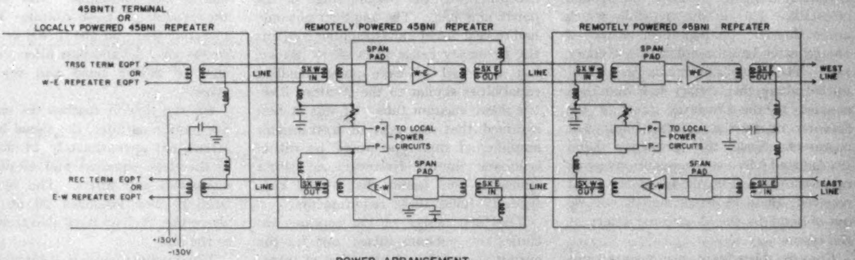
Fig. 5. Block diagram of 45BN repeater


45BNT1 TERMINAL STATION
REMOTELY POWERED 45BN1 REPEATER
TRSG DROP
W
LINE
SX W IN
SPAN PAD
W-E
SX E OUT
WEST LINE
TO LOCAL POWER CIRCUITS
P+
P-
SPAN PAD
REC DROP
E
LINE
SX W OUT
E-W
SX E IN
EAST LINE
-48V
+130V

POWER ARRANGEMENT
TERMINAL FEEDING REMOTELY POWERED REPEATER


REMOTELY POWERED 45BN1 REPEATER
LOCALLY POWERED 45BN1 REPEATER
REMOTELY POWERED 45BN1 REPEATER
WEST LINE
SX W IN
SPAN PAD
W-E
SX E OUT
LINE
TO LOCAL POWER CIRCUITS
P+
P-
EAST LINE
SX W OUT
E-W
SX E IN
-48V
+130V
GRD
+130V
-48V

POWER ARRANGEMENT
LOCALLY POWERED REPEATER FEEDING TWO REMOTELY POWERED REPEATERS


45BNT1 TERMINAL OR LOCALLY POWERED 45BN1 REPEATER
REMOTELY POWERED 45BN1 REPEATER
REMOTELY POWERED 45BN1 REPEATER
TRSG TERM EQPT OR W-E REPEATER EQPT
REC TERM EQPT OR E-W REPEATER EQPT
LINE
SX W IN
SPAN PAD
W-E
SX E OUT
WEST LINE
TO LOCAL POWER CIRCUITS
P+
P-
SX W OUT
E-W
SX E IN
EAST LINE
+130V
-130V

POWER ARRANGEMENT
TERMINAL OR LOCALLY POWERED REPEATER FEEDING TWO
REMOTELY POWERED REPEATERS IN TANDEM

Fig. 6. 45BN repeater power-feed arrangements

amplifier and a vacuum-tube output stage.

For regulation, a portion of the output signal is picked off the output transformer, amplified and rectified, and then passed through a d-c amplifier and time-delay circuit. The output of the d-c amplifier heats a thermistor in the feedback loop of the regulating amplifier. Changes in signal level vary the thermistor resistance to control the feedback loop.

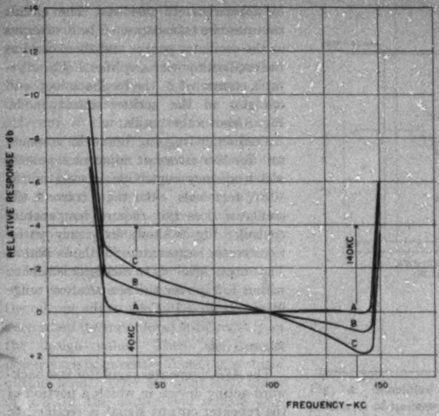
When the repeater is connected for high-low operation, the signal path through the modulator-filter-slope equalizer section is simply reversed. The incoming high group passes through a 164- to 264-kc band-pass filter which removes unwanted low-group crosstalk. This band then modulates the 304-kc signal from the oscillator. The principal outputs are a 40- to 140-kc band and a 468- to 568-kc band. A low-pass filter selects the lower of these. The 40- to 140-kc band undergoes about 2-db negative slope change in the slope equalizer and is then amplified and passed on to the line. The output is readjusted from a +12-dbm (db referred to 1 milliwatt) pilot for the high group to a +3-dbm pilot for the low group. This adjustment is made through a screwdriver-adjusted potentiometer. The potentiometer is easily reached at the front of the unit.

The process for either high-low or low-high operation is identical for both directions of transmission. The amplifier and regulator are flat over the entire band from 40 to 264 kc.

The new repeater meets the frequency response, noise, and intermodulation requirements for the Lenkurt type *45BN* systems as well as for the Western Electric types *N* and *ON* systems. Therefore, the *45BN* repeater will work equally well on any of the three systems.

Power Requirements

The *45BN* repeater requires slightly less than 100 ma (milliamperes) at 120 volts. No separate filament battery is needed. The two 20-volt filaments operate in series across the 120-volt supply. The 80-volt drop supplies part of the voltage required by the transistors



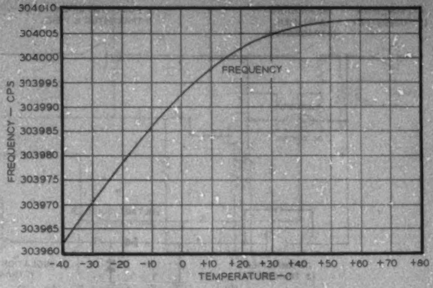


Fig. 9. Oscillator frequency versus temperature

Fig. 7 (left). Slope characteristics for the three settings of the 45BN repeater slope equalizer

so that very little power is wasted. Voltage of either polarity with respect to ground may be used.

Both input and output transformers are tapped with simplex leads. Through these leads, remote d-c power may be fed over the cable pairs used for carrier transmission. Fig. 6 shows several remote power-feed arrangements.

Remote power feed to two adjacent repeaters requires a power supply only at every third repeater. For this arrangement, +130 volts and −48 volts are necessary. These are voltages found in most telephone offices where −48 volts is required for telephone switching and +130 volts is standard for electronic equipment. Each *45BN* repeater has a rheostat to drop the applied voltage to the 120-volt value so that other voltages such as +130 volts and −130 volts may be used for remote power feed.

Even though power is fed over cable pairs whose resistance may vary due to temperature changes, no voltage regulation is required. The repeater tends to be self-regulating. In hot weather, an increase in cable resistance tends to decrease the voltage available at the remote repeater. But the accompanying increase in transmission loss in the cable requires more gain in the repeater. Under this condition, the regulator (and thus the entire repeater) draws less current through the increased cable resistance and the voltage at the repeater remains nearly constant. Under normal conditions, the voltage at the repeater remains constant within ±1 volt.

Power can be fed to two repeaters in tandem by using +130 and −130 volts. Tandem power feed reduces the requirements to only one power supply for every five repeater locations. In this type of operation, the voltage-compensating feature which has been described is not fully achieved because the second repeater draws power through twice the length of cable as before. For this reason, it is recommended that maximum repeater spacing be reduced to 7.8 miles when tandem power feed is used. In addition, the combination of extremely high temperature (such as desert locations) and long system length should be avoided.

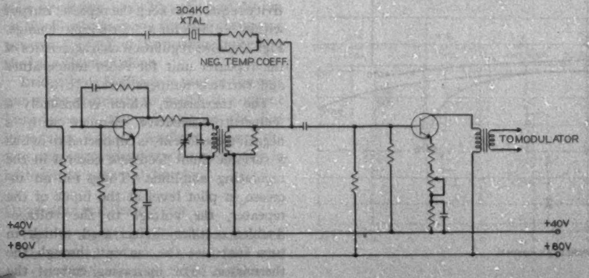


Fig. 8. Oscillator circuit which supplies the 304-kc frogging-frequency to the modulator

Circuit Description

Slope Equalizer

In an 8-mile repeater section, a low group will have about 4 db more slope than a high group. Or, taking the average of the two groups as the nominal slope, the low group has 2 db more than nominal and the high group has 2 db less than nominal. For proper operation of the frequency-frogging process, the two slopes should be equal. Therefore, it is necessary to remove 2 db of slope from received low groups and add 2 db of slope to received high groups. To do this, slope is not added to the received high group but is subtracted from it after it has been translated to a low group. This accomplishes the same purpose and allows a single network to perform both functions.

Since all repeater sections are not of the same length, the received groups will not always have the exact slope stated in the foregoing. To allow for minor deviations, two additional networks are included. These may be connected into the circuit by a screwdriver-operated switch on the face of the repeater unit.

Typical slope characteristics for the three settings are shown in Fig. 7. The *A* position gives a flat slope and the *C*

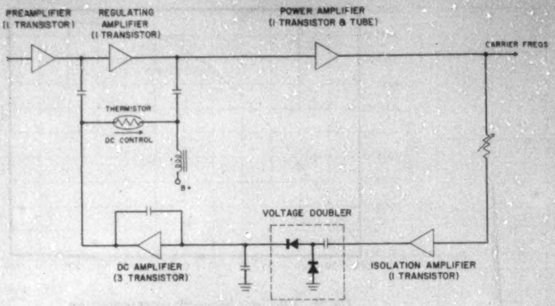


Fig. 10. Block diagram of regulator circuit

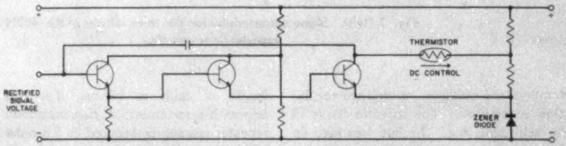


Fig. 11. Schematic of d-c amplifier

position gives −4 db of slope. The normally used *B* position gives −2 db of slope. Thus, adjustments may be made for slope deviations ±2 db from normal. The three positions give equal loss at the pilot frequency so that adjustments may be made without disturbing the output level of the repeater.

When necessary, larger slope deviations are compensated for by artificial lines. A slight negative bulge is introduced into the slope network to cancel the average bulge in the transmission characteristics of each repeater section.

Oscillator

The oscillator supplies the translating or frogging frequency, a nominal frequency of approximately 304 kc. The circuit contains two transistors. The first serves as the oscillator proper and the second as the oscillator power amplifier. The crystal is connected in the feedback loop and operates in the series-resonant mode. Fig. 8 shows the oscillator.

Frequency-frogging tends to compensate for like errors at adjacent repeaters and frequency correction is used in the 45*BN* terminals. For these reasons, the oscillator does not require temperature control. Fig. 9 shows frequency variations versus temperature for the oscillator. The output level of the oscillator is held to within ±5% over the temperature range from −40 to +70 C.

Regulator

The *45BN* repeater regulator is a backward-acting device in which a portion of the repeater output signal is rectified to control a thermistor resistance in the feedback loop of an earlier stage. Fig. 10 is a block diagram of the circuit.

The small portion of the power amplifier output is applied to a voltage-doubler rectifier through an amplifier which isolates the nonlinear rectifier circuit from the output circuit. This prevents the distortion generated in the rectifier from being applied to the line. The output of the rectifier is applied to a 3-stage d-c amplifier which multiplies the current to the proper value for control of the thermistor resistance.

The d-c amplifier consists of two cascaded grounded-collector transistor amplifiers followed by a grounded-emitter amplifier. Fig. 11 is a schematic of this circuit. The capacitor connected in the feedback loop over the three stages provides the necessary time delay in repeater regulation. On the average, the time delay will prevent the repeater gain from changing more than 1 db in the first 10 seconds after signal failure. The zener diode in the emitter lead of the final d-c amplifier provides the stiffness or sensitivity required to keep the repeater output within ±1.5 db for a ±8-db input change. Fig. 12 shows regulation characteristics of the repeater unit for room temperature and extreme temperatures.

The thermistor, which is basically a temperature-sensitive element with a negative coefficient, is connected to act as a variable-shunt feedback control in the regulating amplifier. Thus, for an increase in pilot level at the input of the repeater, the voltage to the voltage-doubler rectifier is increased, which in turn increases the current through the thermistor. For increasing current the thermistor resistance decreases, thus pro-

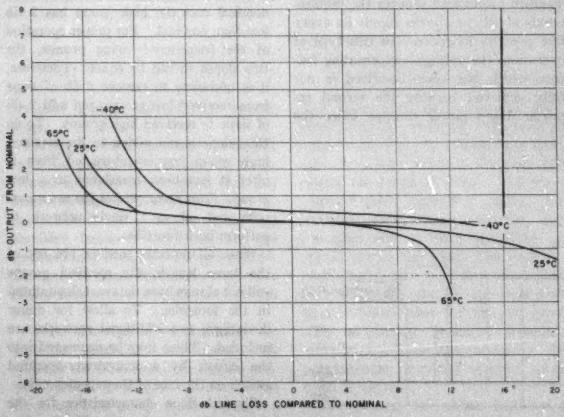


Fig. 12. Regulation characteristics of 45BN repeater for various temperatures

viding more feedback in the regulating amplifier and consequently reducing the gain of the stage.

The regulator circuit is voltage- rather than power-sensitive. This means that the regulator setting will be slightly different if the repeater is used with Western Electric *N* carrier which has 12 pilots, even though the total power of the 12 pilots is equal to the power of the single *45BN* pilot. For this reason, a continuously adjustable control rather than a strapping option has been provided to obtain the different output levels required for high-low or low-high operation.

Further, the continuous control makes the output adjustable to the correct level even when the input level is different from the design value. Thus, the output power variations center about the nominal output level. Over an input level range of ±8 db, the output will remain constant within ±1.5 db at any temperature within the design range of −40 to +60 C.

Power Amplifier

Fig. 13 is a schematic of the power amplifier. The input stage uses a transistor in the grounded-emitter configuration. The transistors in this and the preceding stages that amplify the carrier signals are high-frequency units. These are necessary to obtain wide-band frequency response. The alpha cutoff frequency of these transistors is above 5 megacycles.

Transistors are ideal for low-level stages such as these. Large savings in d-c power are possible. The transistors in these stages operate at a maximum power dissipation of 23 milliwatts. However, in addition to the power dissipated in the transistor, power is required to stabilize the transistor for temperature extremes. In some cases, this will be of the order of the power applied to the transistor itself. However, a comparison of the total power with the power required for a vacuum tube operating with 20 volts and 50 ma on the filament, shows that the transistor power is only 5% of the power required by the tube filament alone.

Bridge-type feedback is used in the output stage primarily to match the output impedance of the repeater to the line impedance. Loop feedback over the two stages provides distortion reduction in addition to gain stabilization.

Over-voltage Protector

Fig. 14 shows a simplified schematic of the method of tandem power feed used for the *45BN* repeater. The resistances represent the d-c resistance of the cable.

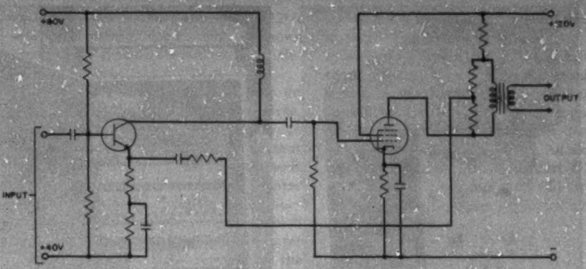


Fig. 13. Schematic of power amplifier

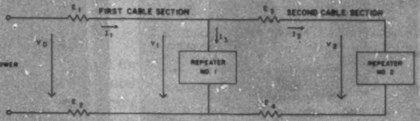


Fig. 14. Simplified schematic of tandem power feed

As can be seen, the voltage at each repeater is dependent on conditions at the other repeater. If a failure occurs at repeater no. 2, so that it draws no current, the voltage at repeater no. 1 may rise to approximately 170 volts. To prevent damage to the repeater unit from such failures, an overvoltage protector is incorporated in the *45BN* repeater.

There are a number of ways in which damaging voltages could occur. The most likely occurrence would be a filament failure or removal of a tube or repeater. A cable break would be another cause. The overvoltage protection circuit is designed to prevent damage to repeaters under these abnormal conditions. Overvoltage protection will not normally be required at a locally powered repeater point.

Fig. 15 is a simplified schematic of the d-c power distribution in the *45BN* repeater. The overvoltage protector consists of a 40-volt zener diode, *CR3*, and its bias resistor, *R4*, and a p-n-p power transistor. The diode provides a reference voltage slightly less than 40 volts that is applied to the transistor base. The transistor emitter is tied to the repeater 40-volt bus and the collector is connected to *B*−.

When the emitter potential is less than 40 volts, the power transistor draws no current, and the overvoltage protector has no effect in the repeater circuit. During abnormal conditions, such as the afore-mentioned, the emitter attempts to rise above 40 volts. It then draws sufficient current to offset the increase in voltage. The overvoltage protector maintains the 40-volt bus within 10% of normal for the most abnormal change. It thus protects transistor circuits and tube filaments from damage by bleeding extra current when the repeater voltage becomes high.

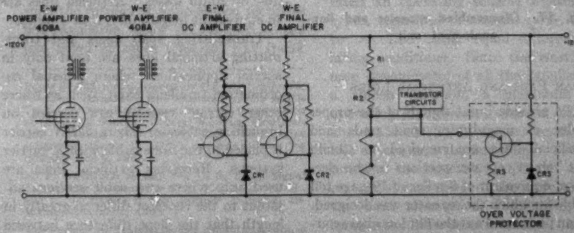


Fig. 15. Simplified schematic of d-c power distribution

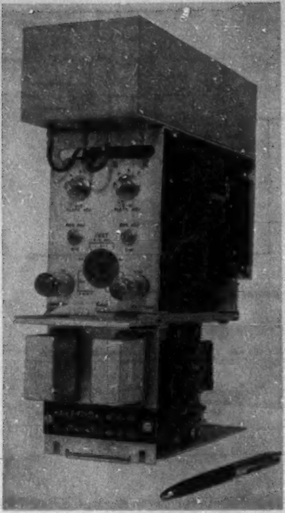


Fig. 16. Assembled repeater and line equipment unit

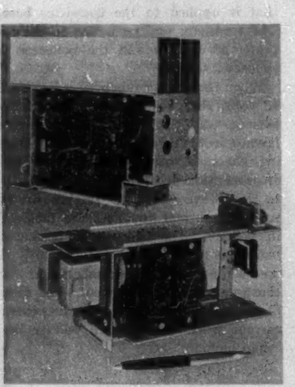


Fig. 17. Disassembled repeater and line-equipment unit

Span Pads and Artificial Lines

To provide input signals of the proper order of magnitude, span pads and artificial lines are required to "build out" short repeater sections to the electrical equivalent of 8 miles of 19-gage toll cable for which the repeater was designed. Span pads build out the flat loss characteristic of the line, and artificial lines build out the slope characteristic.

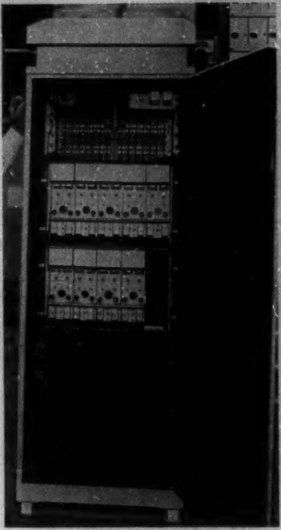


Fig. 18. Nine 45BN repeaters and their line-equipment units installed in cabinet for pole-mounting

The most common requirement is for receiving span pads. Therefore, each repeater is equipped with variable span pads in both recieving branches. These pads may be strapped for any loss from from 0 to 30 db in 2-db steps. The pads are on the equipment side of the line transformer so that their resistance will not decrease the power received over the cable pairs. This is especially valuable when tandem power feed is used.

Although seldom required, space is provided in the repeater for mounting transmitting span pads and both transmitting and receiving artificial lines. Transmitting span pads have 4-, 8-, 12-, 16-, or 20-db losses. Artificial lines are in equivalent lengths of 1, 2, or 4 miles.

Transmitting span pads and transmitting artificial lines are used only in special applications where unusual coordination problems exist, such as where other carrier systems are carried on branching cables or where *45BN* carrier is placed in the same cable with *K* carrier systems. Receiving artificial lines are used only where two cable sections adjacent to the repeater differ so greatly in length that the slope difference between the two cannot be cancelled by the repeater slope adjustment.

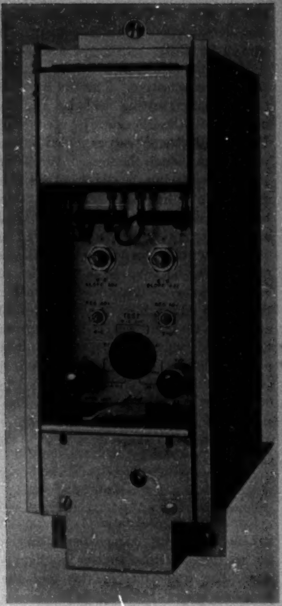


Fig. 19. A 45BN repeater unit mounted in adapter for use in Western Electric type N repeater racks

Mechanical

The complete *45BN* repeater consists of two units: a repeater unit and a line-equipment unit. The line-equipment unit may be replaced with a mounting adapter if the repeater is to mount on Western Electric type *N* repeater shelves. The repeater and line-equipment units are shown assembled and disassembled in Figs. 16 and 17.

The repeater unit contains the oscillator, filters, modulators, slope equalizers, amplifiers, regulator circuits, and transmitting transformers for both directions of transmission. Screwdriver-operated controls for the regulator and slope adjust network appear on the face of the unit. Also on the face is a 11-pin test socket which contains the following bridging test points:

1. The two amplifier outputs.
2. $B+$ and $B-$ leads.
3. The 304-kc oscillator output.
4. The cathodes of the two vacuum tubes.

The repeater unit plugs in and latches to the line-equipment unit or the mounting adapter.

The line-equipment unit contains the receiving span pads and line transformers and a voltage-dropping rheostat and resistor. The overvoltage protector circuit mounts in this unit when required. Space is provided to mount artificial lines and transmitting span pads when these are needed.

The line-equipment unit will contain the span pads and artificial lines which are peculiar to a particular location. Therefore, it is mounted as a semipermanent unit rather than as a latch-in unit. The line-equipment unit is attached to the shelf by two screws at the front of the unit.

Since the outside cable pairs are wired directly to the line-equipment unit, the space at the side of the shelf normally used for a shelf terminal block is available for an additional repeater. Five repeaters and their associated line-equipment units will mount on a single shelf on a 19-inch rack and will occupy only $12\frac{1}{4}$ inches of vertical space. Fig. 18 shows nine *45BN* repeaters installed in a pole-mount cabinet.

When the *45BN* repeater is used on Western Electric type *N* repeater shelves, a mounting adapter is available to take the place of the line-equipment unit. Fig. 19 shows the repeater unit plugged into the adapter. The adapter contains the same circuitry as the line-equipment unit except for circuits which are normally mounted on the *N* repeater shelf itself. The *45BN* repeater in the mounting adapter is mechanically and electrically interchangeable with Western Electric type *N* repeaters on Western Electric rack assemblies.

Conclusions

A new repeater has been developed for cable carrier use. The new repeater takes advantage of the frequency-frogging technique to achieve low-gain requirements and a minimum of external equalization. Transistors have been used extensively to reduce the size and power requirements. The low power requirements, in turn, permit tandem power feed to remote repeaters over the cable pairs. The need for many expensive power sources is thereby eliminated. Although designed primarily for the Lenkurt *45BN* systems, the new repeater will work equally well with Western Electric types *N* and *ON* systems.

References

1. Application and Transmission Features of a New 12-Channel Open-Wire Carrier System, K. E. Appert, R. S. Caruthers, W. S. Chaskin. *AIEE Transactions*, vol. 73, pt. I, Mar. 1954, pp. 18–27.

2. Mechanical Aspects and Component Features of a New 12-Channel Open-Wire Carrier System, A. G. Ewing, F. W. Frazee, Dale Welling. *Ibid.*, pp. 75–81.

3. A New 4-Channel Open-Wire Carrier System Employing Transistors, R. S. Caruthers, W. S. Chaskin, H. K. Krengel. *Ibid.*, vol. 75, pt. I, July 1956, pp. 314–22.

Failure-Rate Studies on Silicon Rectifiers

N. F. BECHTOLD
ASSOCIATE MEMBER AIEE

C. L. HANKS
NONMEMBER AIEE

THE magic word semiconductor has been placed upon the lips of almost every electronically minded individual in the United States through untiring publicity efforts by manufacturers and their representatives. In accomplishing this purpose, the sales promoter often has impaired the future value of these new devices with predictions of unlimited life and hopeful claims for operating capabilities. These claims, although corroborated by theory, apparently have not yet been fulfilled in reasonable quantity with present production techniques for low-power silicon rectifiers. Until greater reproducibility can be attained, derating of both forward current and reverse voltage is necessary for a useful low failure rate.

The data reported herein are the result of only one of many evaluation programs being carried out by the military to determine the suitability of these varied semiconductor devices for the stringent operating environments encountered in tactical military applications. Operating life capabilities of low-power silicon rectifiers are discussed to acquaint systems application engineers with the unexpectedly high failure rates and related reliability factors demonstrated in an evaluation program begun about 18 months ago, and conducted jointly by the U. S. Army Signal Engineering Laboratories and the Battelle Memorial Institute.

Test Program

SAMPLES

Rectifiers used in this evaluation were in the lower power rating range, i.e., less than 500 ma (milliamperes). Their voltage ratings were also low when compared with the ratings of some silicon rectifiers now available, since no rectifier tested was rated at more than 400 PIV (peak inverse volts) nor less than 100 PIV. Rectifiers of higher power ratings were not included because they were still considered to be in the development stage at the outset of this program.

Five manufacturers were represented, and samples tested included both lead- and stud-mounted types of the "top hat" variety. To reduce the variables of extraneous heat-dissipation surfaces during test, the rectifiers were suspended on nonconducting glass filaments. A minimum of 5 inches separated the test samples from each other and from any object that could provide a heat sink. This method of mounting was chosen to eliminate the additional variable introduced by the better dissipation properties of the stud-mounted units.

Differences in forward current rating for the same over-all package size necessitated an operating scheme whereby all types, regardless of manufacturer, would be operated under equal failure-stimulating conditions. Since the least common denominator of all the samples was a silicon junction, a maximum safe operating junction-temperature range of about 150 to 180 C (degrees centigrade) was chosen as a starting point. Average junction temperature is difficult to ascertain, so case temperature was chosen as the next alternative, recognizing that some minor differences in internal package design could result in some inequalities in heat-dissipation properties. These

---

Paper 57-795, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing July 9, 1957.

N. F. Bechtold is with the U. S. Army Signal Engineering Laboratories, Fort Monmouth, N. J., and C. L. Hanks is with Battelle Memorial Institute, Columbus, Ohio.

The authors are grateful for the helpful suggestions and services provided by K. E. Hassler and A. Mace of Battelle Memorial Institute, and S. Danko of the U. S. Army Signal Engineering Laboratories.

Table I. Rated and Actual Direct Currents for Load-Life Conditions

Manufacturers Code	Manufacturers Rating: Output Current, Ma	Manufacturers Rating: PIV, Volts	Forward Current Required for Specific Case-Temperature Rise, Ma: At 40 C	Forward Current Required for Specific Case-Temperature Rise, Ma: At 20 C
BU	200	100	400	220
	200	400	400	275
	400	100	500	350
	400	400	500	375
BV	200	100	500	300
	200	350	500	150
BW	300	100	400	250
	300	400	400	275
BX	250	100		350
	250	200		375
BY	300	100	425	
	300	400	400	

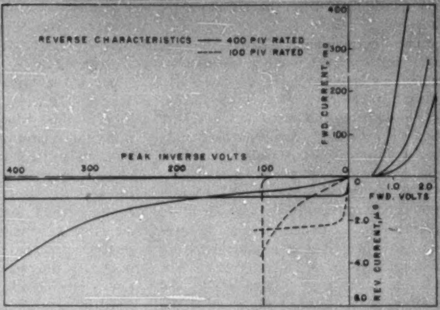


Fig. 2. Spread of initial characteristics at 100 C

design differences can be attributed to either thickness and surface area of the silicon wafer, or fabrication of the junction proper. With regard to the latter technique, some manufacturers use the circular area of a wire butted against the silicon surface, while others achieve a larger junction area by using a greater portion of the silicon wafer cross section.

Test Procedure

Initial tests to determine the output current required to produce a desired temperature rise at the base of the rectifiers were performed on several units of each type and manufacture. The results shown in Table I are the approximate averages of the currents required. The 40 C rise test was specifically designed to stimulate failure, since it was recognized that forward currents in excess of manufacturers' ratings, necessary to achieve this condition, would contribute to significant junction heat rise. The data gathered for these conditions were used merely as a basis for study of failure mechanisms, and not considered as part of an over-all quality evaluation. The 20 C temperature rise, however, required that test samples be operated at or near the manufacturer's specification, thereby simulating normal-rated operation. Each initial sampling test established an operating current for each manufacturer's product in a subsequent related load-life test.

Three load-life tests were conducted during this program as follows: 1. an initial case-temperature rise of 40 C at the base of the rectifiers in an ambient temperature of 100 C, 2. an initial case-temperature rise of 40 C in an ambient temperature of 25 C, and 3. an initial case-temperature rise of 20 C in an ambient temperature 100 C. With one exception, groups of all the rectifier types and manufacture in a particular test were operated at their rated voltage. To study the effect on failure rate of reverse voltage alone, groups of the low-current higher-voltage (400 volts) types of manufacturers *BU*, *BV*, and *BW* were operated with resistive-capacitive loads at 75% of their rated voltage in the 25 C and 100 C ambient load-life tests with a 40 C initial case-temperature rise.

Load was applied continuously to the rectifiers during these tests, and measurements made with a dynamic analyzer. Fig. 1 is a circuit diagram of this test unit. The circuit employs three blocking diodes which enable the operator to measure the alternating voltage applied, output voltage, output current, dynamic forward voltage drop, and dynamic reverse leakage current simultaneously, without interruption of load. In addition, thermocouples attached to each individual recti-

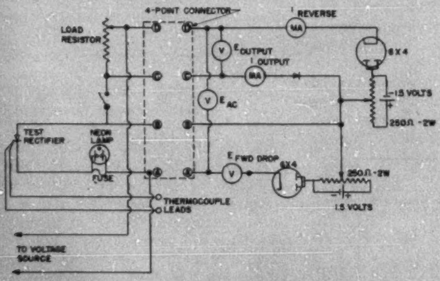


Fig. 1. Dynamic analyzer circuit for load-life tests

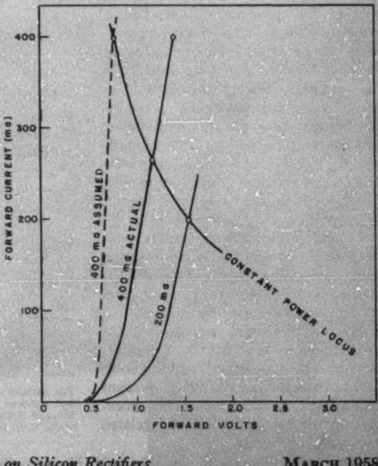


Fig. 3 (right). Forward characteristics and power generated

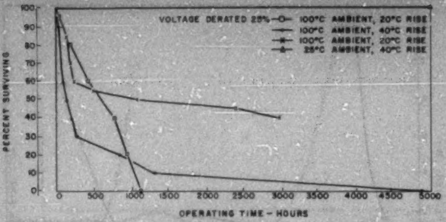


Fig. 4. Survival of 400-ma 400-PIV silicon rectifier, BU manufacturer

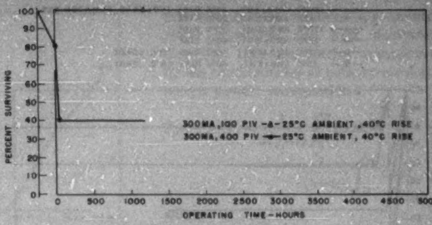


Fig. 6. Survival of 300-ma 100-PIV, and 300-ma 400-PIV silicon rectifier, BY manufacturer

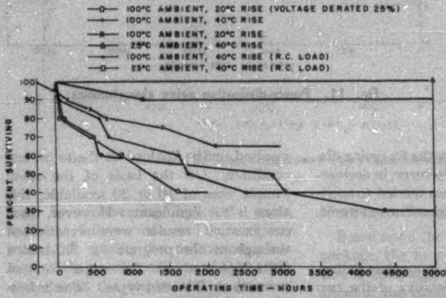


Fig. 5. Survival of 200-ma 400-PIV silicon rectifier, BU manufacturer

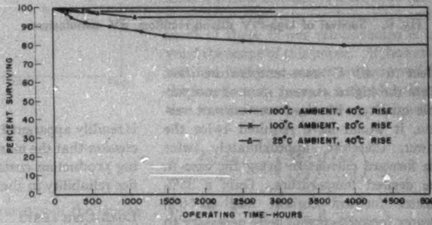


Fig. 7. Survival of low-PIV silicon rectifiers

fier were used to measure base temperature of each rectifier.

For clarity and simplicity, the following symbolic notations are used hereafter to indicate the specific related test conditions or ratings as noted.

Notation	Test Condition or Rating
100+40...	100 C ambient operating temperature, 40 C case-temperature rise
100+20...	100 C ambient operating temperature, 20 C case-temperature rise
25+40...	25 C ambient operating temperature, 40 C case-temperature rise
100−*R*...	rated at 100 PIV, operated into resistive load
400−*R*...	rated at 400 PIV, operated into resistive load
400−*RC*..	rated at 400 PIV, operated into resistive-capacitive load
a/*b*.......	*a* is the load current value at which a unit is operated for a specific condition, and *b* is rated load current of that unit as specified by the manufacturer

Example: *BU* (400/200), (100+40) means the 200-ma rated item from manufacturer *BU* was operated at 400 ma to generate a 40 C case-temperature rise in a 100 C ambient.

Discussion of Results

PRELIMINARY EVALUATION

As indicated in the foregoing reference to Table I, preliminary tests were run on small sample sizes in order to establish basic operating conditions for load-life tests. At this early stage, these samples were assumed to be representative of the larger batches to follow. Minor variations of ±5 C were, at the time, satisfactory justification for establishing the test points.

Initial characteristic tests performed thereafter on the main body of samples gave indication that the foregoing reasoning was fallacious. Fig. 2 demonstrates the initial tolerances exhibited in the products of one manufacturer; for any test performed close to a critical operating point, it is apparent that there would be many deviations from an operating temperature established as an average from a small sample. Since forward power losses initially contributed to more than 90% of the total dissipation, it is evident that small variations in forward voltage drop can result in very significant differences in current necessary to attain a specified case temperature. In this regard, the item having an extremely low forward drop can be greatly penalized by a necessary high current density which, although resulting in required case temperature, has provoked a junction hot-spot temperature beyond the safe limit. (As is shown later, variations between reverse characteristics also become significant after some aging has taken place.)

Consider, for instance, samples from manufacturer *BU*, rated basically at 400 ma and 200 ma, as illustrated in Fig. 3. The 100-PIV versions of this series required 500 and 400 ma, respectively, to

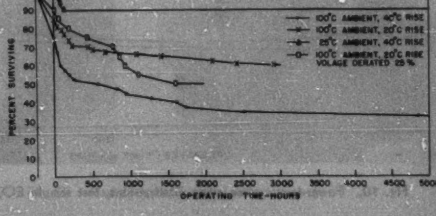


Fig. 8. Survival of high-PIV silicon rectifiers

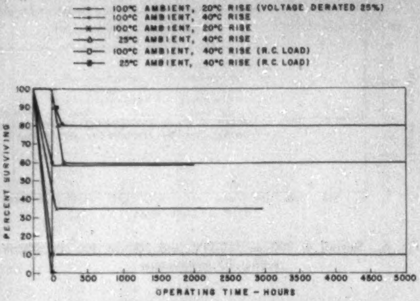


Fig. 9. Survival of high-PIV silicon rectifiers, BV manufacturer

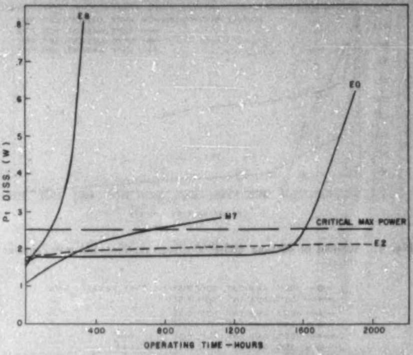


Fig. 11. Power-dissipation aging characteristics

effect a 40 C case-temperature rise. Were the higher current item of comparable quality with the lower current version, it would have required twice the power, therefore, approximately twice the forward current to bring the case to the desired temperature. Only a 25% greater forward current and a 50% greater forward current were necessary to bring the 400-ma unit to case-temperature rises of 40 C and 20 C respectively as shown in Table I. Even these differences are easily explained, since reverse current increases and contributes more to heating with increased forward currents and junction temperatures. Were the 400-ma curve to fall near the dashed line of Fig. 3, the desired result would be achieved since each would generate equal power at its respective rated value. It is readily apparent from the foregoing discussion that the manufacturer, in decreasing production costs, has tended to sacrifice reliability in the higher current items.

LOAD-LIFE TESTS

Figs. 4 and 5 give an immediate indication of the relative reliability of the two types discussed in the foregoing. Suggestions are made later for possible failure-rate improvement of the 200-ma type, but the 400-ma version is apparently below even the lowest military quality standards in the 100 C ambient operating range. Note from Fig. 5 that no failures were noted under the (25+40) condition *R* load, whereas 10% failures resulted under *RC* load at the same test condition. On the basis of the small sample sizes of 10 or 20 available, this alone is not significant. However, these comparative results were experienced throughout the program for *RC*-loaded units operated at the same forward currents as *R*-loaded types. The manufacturer's recommended 20% derating for *RC* load appears adequate in the light of these results.

In view of excessive failures of higher voltage (400 PIV) units during the first 1,000 hours, it was decided to study the effect on failure rate of reverse voltage alone. Consequently, another test run was initiated at the (100+20) condition, with reverse voltage derated to 75%. Fig. 4 demonstrates an apparent advantage gained from this approach but reference to Fig. 5 apparently indicates a decrease in quality. These inconsistencies are a result of small sample size, and further tests must be run on larger sample sizes to determine the relative effects of various voltage-derating factors. Fig. 6, representing the comparative survival rates of low- and high-PIV items of manufacturer *BY*, indicates clearly the differences in quality at the particular test condition. Figs. 7 and 8 demonstrate load-life results of the two categories (100−*R*) and (400−*R*) respectively. The curves of Figs. 7 for (100−*R*) denote the closest approximation to desired military quality, especially for the (100+20) condition, which indicates a failure rate of less than 1% per 1,000 hours. Derating currentwise of even this series of items would be necessary to decrease this failure rate to a reasonable figure in the neighborhood of 0.1% per 1,000 hours.

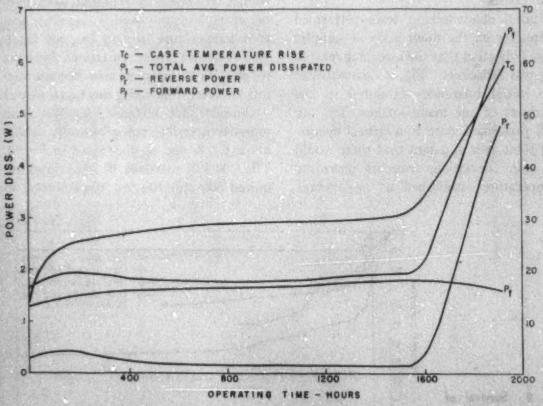


Fig. 10. Power-temperature aging relationships, test sample EO, BU manufacturer

Vol.
Pt. 1
1958

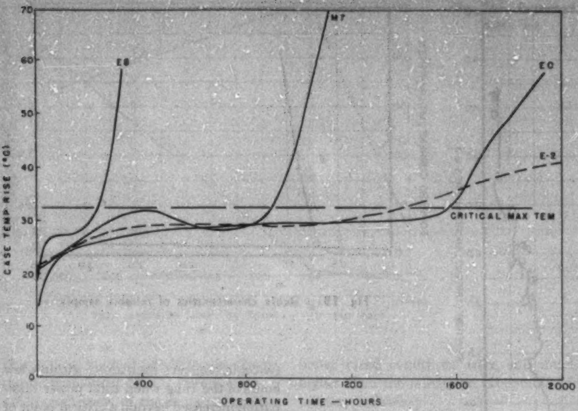


Fig. 12. Case-temperature aging characteristics

Fig. 8, representing high-PIV units, indicates failure rates well beyond any reasonable quality for military application at the rated condition of (100+20), and also at the accelerated condition of (25+40). It appeared throughout this program that the 400-volt items were more affected by high ambient temperature rather than by operating currents in excess of rated values.

Each of the foregoing discussed survival curves (and therefore failure-rate curves) approximated a decreasing exponential function of $y=e^{-as}$, or a constant "hazard" rate (per-cent failure of remaining samples), resulting in an early high per cent of failure which decreases with time. This effect is accentuated in Fig. 9, which indicates no failures for manufacturer *BV* samples beyond an initial burn-out period of about 200 hours. At an increase in unit cost, the manufacturer could weed out these undesirables with a 1-week factory test prior to shipment. Consider the effects of such a burn-out period from Figs. 7 and 8 and it is evident that the advantage gained would be significant in almost every case.

Table II. Modes of Failure

Effect	Cause
Overheating	overload, high ambient, poor dissipation
Junction breakdown	high reverse voltage
Loss of seal	poor quality control
Microphonics	loose elements—initial or after vibration
Insulation failure	moisture absorption, high potentials, poor basic design

NATURE OF FAILURES

Based upon the geometry of the device package, the heat generated at the junction is dissipated through thermal resistances of the junction, internal mounting members, and case. As previously stated, no external heat-dissipation surfaces were considered. Consequently, relatively high thermal resistance of the small devices (exceeding 0.1 C per megawatt) resulted in average junction temperatures higher than case temperature, hot spots considerably higher than this, and a purposely induced high failure rate for study.

The operating characteristics discussed here are all dependent upon the basic time-temperature relationships between power generated and heat dissipated. However, they are further complicated by the contributions of both forward and reverse power; see equation 1.

$$P_t=P_f+P_r=\frac{1}{(t_2-t_1)_f}\int_{t_1}^{t_2} e_f i_f dt+\frac{1}{(t_2-t_1)_r}\int_{t_1}^{t_2} e_r i_r dt \quad (1)$$

where

$(t_2-t_1)_f$ = period of flow of forward current
$(t_2-t_1)_r$ = period of flow of reverse current

In view of the nonlinear forward and reverse characteristics of a semiconductor rectifier, a complete mathematical treatment of these power relationships is beyond the scope of this paper. If, however, rectified half sine waves are assumed for the functions e_f, i_f, e_r, i_r, a simple calculation results in average power relationships of

$$P_f=\frac{E_{Mf}I_{Mf}}{2} \text{ or } \left(\frac{E_{Mf}}{\sqrt{2}}\right)\left(\frac{I_{Mf}}{\sqrt{2}}\right) \quad (2)$$

$$P_r=\frac{E_{Mr}I_{Mr}}{2} \text{ or } \left(\frac{E_{Mr}}{\sqrt{2}}\right)\left(\frac{I_{Mr}}{\sqrt{2}}\right) \quad (3)$$

where

E_{Mf} = peak value of forward voltage drop
I_{Mf} = peak value of load current
E_{Mr} = peak value of input voltage
I_{Mr} = peak value of reverse leakage current

Equations 2 and 3 signify that average power for each half-cycle is merely the rms products of voltage and current present during that half-cycle. These relationships were used to approximate magnitudes for the following discussion.

The actual contributions of both P_f

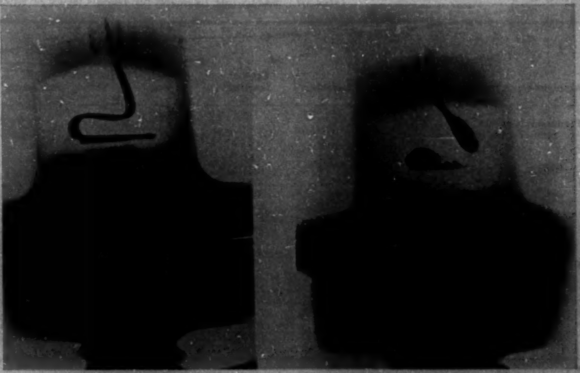


Fig. 13. Radiographs of typical failures

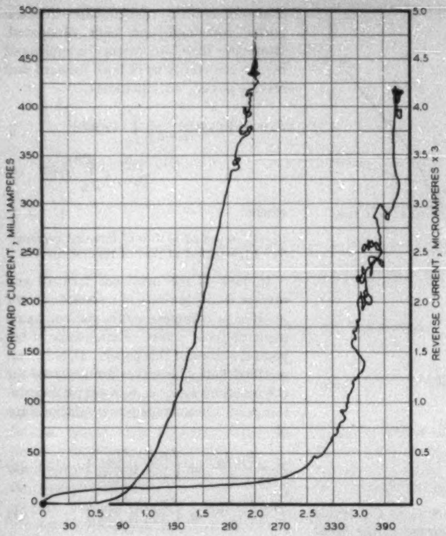


Fig. 14 (left). Forward and reverse instability

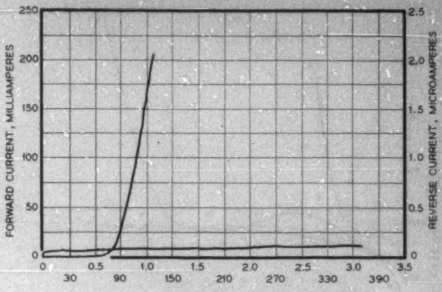


Fig. 15. Stable characteristics of reliable sample

and P_r to ultimate failure are illustrated in Fig. 10, where unit *E0* has been chosen as a typical example. The power dissipated due to forward losses remains essentially constant, at approximately 90% of the total losses, throughout the 1,500-hour useful life of the sample. Almost simultaneous runaway of reverse power, total power, and case temperature occur beyond this point. Forward and reverse power losses contribute equally at about 1,700 hours under this particular test condition. Figs. 11 and 12 further illustrate this temperature-power relationship for three other samples in addition to unit *E0*. *E8* and *M7* were early failures, while *E2* was still operating satisfactorily after 2,500 hours. It is especially significant that *E0*, *E8*, and *M7* reached a critical case-temperature rise of about 33 C (without recovery thereafter) within a few hours of the time when total power dissipation increased beyond a critical point of about 250 megawatts, thereby automatically limiting this series to about 400 ma d-c with negligible reverse losses. Further, with regard to Fig. 12, case temperature apparently rises rapidly during a 100-hour break-in period, then levels off for a period of time determined by end quality of the individual sample. At the point of permanent change, a high irreversible rate of temperature increase (in excess of 10 C per 100 hours) takes place to precede thermal runaway. Unit *E2*, however, did not experience these effects during the test period, and has provided satisfactory operation for at least 2,500 hours. Apparently some deterioration has taken place in *E2*, as evidenced by increase in case temperature beyond 33 C.

Table III. Mean-Life Estimates for Three Classes of Silicon Rectifiers

Ambient Conditions	Class of Rectifier	Type of Loading	Duration of Test, Hours	Reliability Data: Sample Size	Reliability Data: Successes	Reliability Data: Failures	Mean Life, Hours: Average Estimate	Mean Life, Hours: 95% Lower Confidence Limit	Approximate Failure Rate Per 1,000 Hours, Per Cent
Planned initial 40 C base temperature rise									
25 C	100 PIV	*RES**	5,009	40	38	2	95,982	40,456	1
	400 PIV	*RES*	5,009	30	30	0	150,270	50,174	†
	400 PIV	*RC*‡	5,009	20	9	11	5,875	3,813	17
	350/400 PIV	*RES*	5,009	40	36	4	45,148	23,302	2
	350/400 PIV	*RC*	5,009	30	17	13	8,066	5,391	12
100 C	100 PIV	*RES*	5,013	40	32	8	21,339	12,982	5
	400 PIV	*RES*	5,013	30	13	17	4,701	3,289	21
	400 PIV	*RC*	5,013	20	10	10	6,123	3,900	16
	350/400 PIV	*RES*	5,013	40	13	27	2,960	2,214	34
	350/400 PIV	*RC*	5,013	30	11	19	3,486	2,481	29
Planned initial 20 C base temperature rise									
100 C	100 PIV	*RES*	2,925	100	98	2	143,688	60,564	0.7
	200 PIV	*RES*	2,925	20	20	0	58,500	19,533	†
	400 PIV	*RES*	2,925	60	41	19	7,067	5,029	14
	350/400 PIV	*RES*	2,925	80	48	32	4,838	3,699	21
100 C, 25% derated	400 PIV	*RES*	1,945	15	7	8	2,414	1,469	41
	350/400 PIV	*RES*	1,945	20	10	10	2,515	1,602	40

* Resistive load.

† Sample size too small for estimation of reasonable population failure rate greater than 95% of time.

‡ Resistive-capacitive load.

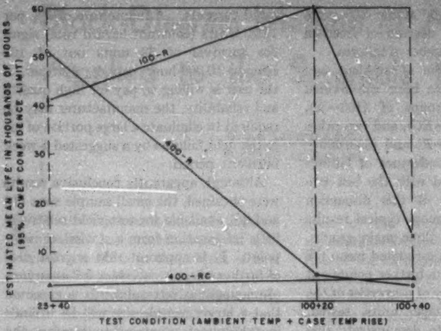


Fig. 16. **Estimated mean life**

This was a gradual gradient, however, and no permanent defects were noted.

Although some of the units under discussion were tested under accelerated operating conditions, the trends and overall results are indicative of, but more pronounced than, normal operation. However, the failure rates experienced under rated conditions were still much higher than the maximum acceptable for military reliability.

It is logical to conclude from the foregoing analysis that silicon rectifiers of the low-power types studied give prior warning of ultimate failure through a measureable high rate of change of case-temperature rise and power dissipation. Ultimate failure and subsequent lack of equipment reliability can therefore be avoided by close monitoring of diode characteristics, and replacement within 100 hours of any items having exceeded the critical points specified. A more practical alternative to the extension of equipment reliability is to provide sufficient derating of forward and reverse operating parameters, so as to extend these points of failure and to eliminate this need for component monitoring during equipment operation. Specific derating parameters are discussed later.

MECHANISMS OF FAILURE

For the purposes of this discussion, "end of life" is defined as that point where either the rectifier ceases operation catastrophically or thermal runaway is taking place. A maximum reverse leakage of 125 ma was chosen as the criterion for the latter condition.

Although test samples were essentially of equal mechanical and physical construction, failures noted were of both the

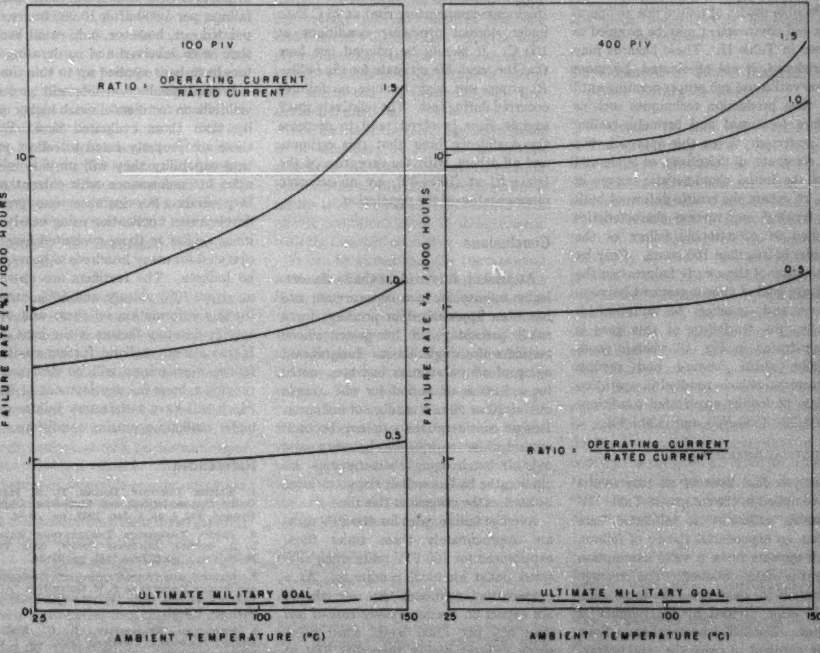


Fig. 17. **Proposed failure rate curves, low-power silicon rectifiers**

open- and short-circuit variety. The open-circuit-type failure was unique in that it only occurred with rectifiers of manufacture *BU*. This type of failure occurred in all three load-life evaluations and was generally preceded by an increasing forward voltage drop and temperature rise at the base of the rectifier. The reverse leakage current varied generally, with increases to several milliamperes in a few cases. Radiographs of two rectifiers that suffered this type of failure are shown in Fig. 13. Rectifier *A* of this figure has a separation at the alloyed area between the silicon wafer and the aluminum wire. Internal arcing has resulted in the melting of the aluminum wire in rectifier *B*. The failure-by-short-circuit effect was general in the higher PIV rectifiers of all the manufacturers. This type of failure follows a gradually increasing reverse leakage current that sometimes reaches several milliamperes, which then increases to a value greater than 125 ma or approaches a direct short circuit (thermal runaway). Actual short-circuiting was confined to rectifiers on resistive-capacitive load in the (100+40 C) and (25+40 C) load-life evaluations.

Possible modes of failure due to operation and environment may be grouped as shown in Table II. These hazards may be reduced, if not eliminated, by more conservative use and proper derating until the time production techniques such as surface treatment and hermetic sealing are perfected. Even this approach will not eliminate all defectives, as evidenced from the initial characteristic curves in Fig. 14, where the erratic nature of both the forward and reverse characteristics resulted in catastrophic failure of the rectifier in less than 100 hours. Only by elimination of these early failures can the ultimate goal of close agreement between theory and practice be approached. Ideally, the feasibility of this goal is demonstrated in Fig. 15, wherein good, initially stable, forward and reverse characteristics have resulted in operation to date under the accelerated conditions of 500/200, (350-*RC*), and (100+40).

STATISTICAL ANALYSIS

Analysis had been of all test results and compiled in the composite Table III. Mean-life estimates as tabulated here assume an exponential theory of failure, which appears to be a valid assumption from the data obtained. The average estimate is the quotient of the total hours of operation divided by the number of failures. The 95% lower confidence limit is the quotient of twice the total operating hours divided by χ^2 at the 0.05 probability level with degrees of freedom equal to twice the number of failures.

Each test condition of ambient and case temperature has been sub-divided into three main groups of (100−*R*), (400−*R*), and (400−*RC*), and two other groups of (350/400−*R*) and (350/400−*RC*). Since a preponderance of failures has been experienced with the last two categories, the rest of this discussion relates only to the more typical results noted for the other three main groups. A consistently longer estimated mean life for the (100−R) units further points out the relative inequality of properties of the low- and high-voltage units tested. Fig. 16 may tend to demonstrate this point. (Statistically, it is incorrect to connect the points into a continuous function, since different sample size and test periods were involved at each temperature condition. This approach was taken only for the purposes of text clarification.) Note especially how the 400-volt units, under both loading conditions, suffered greatly at the higher ambient temperatures, while the 100-volt items were more adversely affected by high current density (high case-temperature rise) at 25 C than under normal operating conditions at 100 C. It should be pointed out here that the mean-life estimate for the (400−*R*) group was high because no failures occurred during test. The relatively small sample sizes involved tend to decrease the confidence limit that this estimate and all others, with the exception of the (100−*R*) at (100+20), are an accurate representation of the population.

Conclusions

At present, it is evident that a theoretically infinite life can be expected, and has been approached in practice, for a small percentage of low-power silicon rectifiers of a single batch. Insignificant aging of all parameters has been noted for a 5,000-hour period for one sample out of 10 or 20 in a number of instances. Design considerations can not be built around these exceptions, however, and over-all batch aging characteristics, including the bad as well as the good, must be used as the criteria at this time.

Average failure rates for 400-PIV units are approximately three times those experienced for 100-PIV units when analyzed under identical conditions. As a conservative estimate, the user should not expect an average failure rate of less than 5% per 1,000 hours when using single cells at higher than 200 PIV at rated currents. A failure rate of 5% per 1,000 hours (constant hazard rate) signifies survival of 62 units out of 100 after a 10,000-hour period. Further, if the user is willing to pay for high quality and reliability, the manufacturer may be required to eliminate a large portion of his initial field failures by a suggested 1-week burn-out period.

Although apparently conclusive results were obtained, the small sample sizes (10 and 20) available for test yield relatively little information from a statistical viewpoint. It is apparent that a great deal of further effort is necessary for assurance of unquestionable results. It is expected that a larger sample size will be utilized in subsequent programs to increase the level of confidence and provide sufficient information to complete the curves noted in Fig. 17. Three points on each set of curves have already been approximated from the test results to date and may serve as a guide toward conservative design and application. A 100-PIV unit at rated current, for example, when operated in a 100 C ambient, will yield almost seven failures per 1,000 in the first 1,000 hours, and more than 60 failures per 1,000 after 10,000 hours. As pointed out, however, only small sample sizes were involved and no derating factors have been applied up to this time.

Present production lots will probably yield silicon rectifiers of much higher quality than those evaluated here. When these are properly rated to reflect practical capability they will provide failure more in conformance with current military needs. For instance, one present development application using nearly 100 items similar to those evaluated here has operated for many hundreds of hours with no failures. The rectifiers are operated at about 70% voltage and 50% current. Further information on these and other specific derating factors is the next goal. If they are not realistic, further studies of failure mechanisms will be designed to provide a basis for development of items which will have satisfactory failure rates under realistic operating conditions.

References

1. SILICON JUNCTION DIODES, D. E. Mason. *British Communications and Electronics*, London. England, vol. 4, no. 1, Jan. 1957, pp. 10-14.
2. POWER TRANSISTOR TEMPERATURE RATING, L. A. Griffith. *Electronic Design*, New York, N. Y., vol. 3, no. 3, June 1955, pp. 22-24.
3. RATING AND APPLICATION OF GERMANIUM AND SILICON RECTIFIERS, F. W. Gutzwiller. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 753-57.
4. PREDICTED RELIABILITY, J. C. Kinkaid. *Military Electronics*, Cleveland, Ohio, vol. I, Oct. 1956, pp. 32-35.

Block-Diagram Method of Analysis Applied to the Saturable Reactor

R. M. HUBBARD
ASSOCIATE MEMBER AIEE

THE saturable-reactor class of magnetic amplifiers has recently received considerable attention as the subject of new analyses efforts.[1,2] Although the saturable reactor antedates the self-saturating class of magnetic amplifiers, it is by no means functionally simpler and does not afford the easiest starting point for a general analysis of magnetic amplifiers. Therefore, in a previous paper,[3] the author introduced a block-diagram method of analysis as applied to the self-saturating class of magnetic amplifiers. It was shown that both half-cycle response circuits and conventional full-wave circuits can be analyzed by a single theory and that, in fact, the two differ only by the amount of feedback present. As an extension of the block-diagram method of analysis to the saturable reactor, this paper is intended to lend further support to the premise that all classes of magnetic amplifiers are basically of the same genus and therefore should be describable by a single theory. The use of the block diagram is particularly effective in clarifying the relationship between the saturable-reactor class and the self-saturating class of magnetic amplifiers.

Philosophy of Saturable-Reactor Analysis

The basic mode of operation of saturable reactors is such that, for purposes of a general transfer function, consideration of increments of time smaller than one half-cycle of the carrier frequency is meaningless. The operation of saturable reactors can, however, be described in terms of half-cycle average values. Although analyses based on instantaneous values provide a valuable insight into the mechanics of saturable-reactor operation, the details of the various operating modes tend to detract from a clear understanding of the basic nature of the device.

The simplicity of the following analysis is dependent on the fact that saturable-reactor operation can be examined over discrete (half-cycle) increments of time if any commutation effects are negligible. Therefore, it is assumed that the load current goes to zero at the end of each half-cycle of the carrier frequency. This assumption effectively restricts the operation of the saturable reactor to a single mode and, in fact, is substantially the same assumption previously used to simplify the analysis of the self-saturating class of amplifiers. Fortunately, the resultant general conclusions which are reached with the help of this assumed condition are not invalidated by its absence. Several operating conditions inherently satisfy the requirements of this assumption. For example, the parallel saturable reactor always operates in the assumed manner. In addition, in the case of the series saturable reactor, the use of a square-wave supply voltage or a rectified a-c control voltage with a sinusoidal supply voltage normally assures that the load current will go to zero when the supply voltage changes sign. Clearly, highly inductive loads produce a type of operation which can not be handled by this method of analysis.

The fact that the foregoing assumption is of secondary importance is further evidenced by the excellent correlation between the theory and the experimental results (Fig. 17) which were obtained with circuit conditions often in disagreement with the assumed condition.

It can be shown that the conventional methods of circuit analysis (Kirchhoff's laws, etc.) apply to the analysis of magnetic-amplifier circuits when all voltages are expressed in terms of half-cycle average values as well as when all voltages are expressed in terms of instantaneous values. The use of half-cycle average values, however, has the advantage that no consideration need be given to the various changes of state resulting from core saturation. Therefore, the only necessary equivalent circuit is merely the original circuit diagram with all voltages expressed as average values

Nomenclature

A = amplification factor
A_i = intrinsic amplification factor due to core characteristic
A_i' = intrinsic amplification factor due to coupling between reset and load circuits
A_o = generalized amplification factor
$\bar{e}_c$ = half-cycle average control voltage
$\bar{E}_c$ = average steady-state control voltage
$\bar{e}_g$ = half-cycle average gate voltage across load winding
$\bar{e}_g'$ = half-cycle average gate voltage referred to control winding
$\bar{e}_L$ = half-cycle average load voltage (across R_L)
$\bar{e}_o$ = half-cycle average output voltage
$\bar{E}_o$ = average steady-state output voltage
$\bar{e}_r$ = half-cycle average reset voltage across load winding
$\bar{e}_r'$ = half-cycle average reset voltage referred to control winding
$\bar{E}_s$ = average rectified supply voltage
$G(s)$ = frequency variant portion of transfer function $KG(s)$
$\bar{I}_c$ = half-cycle average current in control winding
K = static voltage gain associated with transfer function $KG(s)$
N_c = number of turns in control winding
N_x = number of turns in load winding
$\bar{R}$ = average excitation impedance of saturable transformer referred to the load winding
R_c = total series resistance in control circuit
r_c = resistance of control (N_c) windings
R_L = load resistance
R_r = leakage resistance of the rectifier
r_x = resistance of load winding plus the forward resistance of the rectifier
s = complex frequency variable
T = time for one half-cycle of supply (ω_c)
τ_a = amplifier 63% response time
ω_c = supply frequency in radians per second
ω_0 = "break frequency" or frequency for 45-degree phase shift

Development of Block Diagram for Series Saturable Reactor

Reference to the series saturable-reactor circuit diagram of Fig. 1(A) permits a simple introduction to the block-diagram type of analysis. The two component saturable transformers T_1 and T_2 are represented in equivalent-circuit form in accordance with the technique previously applied to the self-saturating circuits.[3] Thus, the nonlinear excitation impedance is represented by an average resistance $\bar{R}$ and the resistance of each load winding is lumped into the total series load-winding resistance r_x. Likewise the resistance of each control winding is lumped into the total series resistance of the control circuit R_c. It should be noted that the representation of the nonlinear magnetic characteristics of the core by the parameter $\bar{R}$ is not an attempt to linearize this quantity. By including this one nonlinear factor in the block diagram without specifically evaluating

---

Paper 57-923, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 27, 1957; made available for printing July 29, 1957.

R. M. Hubbard is with Boeing Airplane Company, Seattle, Wash.

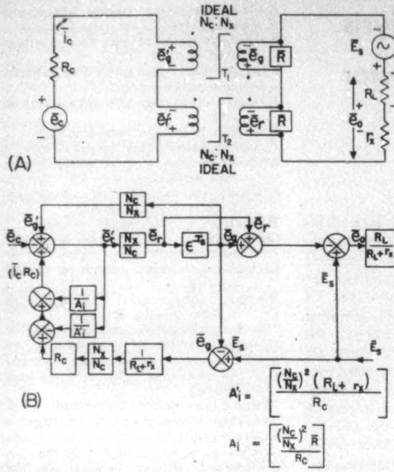

IDEAL
$N_c:N_x$
T_1
T_2
$\bar{R}$
E_s
R_L
$\bar{e}_L$
$\bar{e}_o$
r_x
R_c
$\bar{e}_c$
(A)
(B)
$A_i' = \left[\dfrac{(N_c/N_x)^2 (R_L + r_x)}{R_c}\right]$
$A_i = \left[\dfrac{(N_c/N_x)^2 \bar{R}}{R_c}\right]$

Fig. 1. Series saturable reactor

A—Circuit diagram with component saturable transformers represented in equivalent-circuit form

B—Block diagram

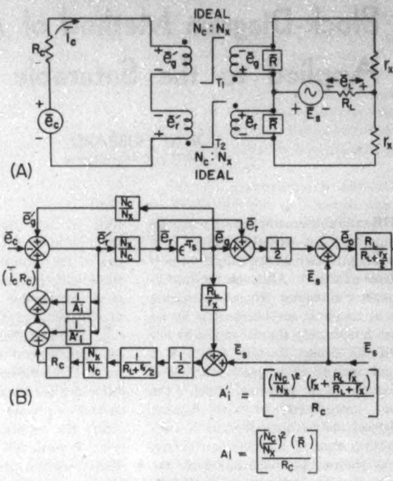

IDEAL
$N_c:N_x$
T_1
T_2
$\bar{R}$
r_x
R_L
$\bar{e}_L$
E_s
R_c
$\bar{e}_c$
(A)
(B)
$A_i' = \left[\dfrac{(N_c/N_x)^2 \left(r_x + \frac{R_L r_x}{R_L + r_x}\right)}{R_c}\right]$
$A_i = \left[\dfrac{(N_c/N_x)^2 (\bar{R})}{R_c}\right]$

Fig. 3. Parallel saturable reactor

A—Circuit diagram

B—Block diagram

it, important general conclusions can be easily reached. The evaluation of $\bar{R}$ is rightfully the subject of a separate investigation such as the very excellent work recently described by Kikuchi and Feth.[4,5]

For a given saturable transformer, the net flux change over a complete cycle of resetting and subsequent gating must be zero. In addition, the application of the previously mentioned assumption effectively assures a symmetrical type of operation such that the magnitudes of the flux changes over the two equal half-cycle increments of time are equal. Therefore, the magnitude of the volt-time integral absorbed by the load winding N_x during the resetting half-cycle equals the magnitude of the volt-time integral absorbed by the same winding during the subsequent gating half-cycle. Thus, in the case of a given transformer, the magnitude of the average gate voltage in the nth half-cycle $\bar{e}_g(nT)$ equals the magnitude of the average reset voltage in the preceding $(n-1)$th half-cycle $\bar{e}_r(n-1)T$.

The following description of the operation of the series saturable reactor of Fig. 1(A), based on the block diagram of Fig. 1(B) and the transformed block diagram of Fig. 2, serves to illustrate the straightforward relationship between circuit diagram and block diagram. An analytical development of the block diagram is presented in Appendix I.

During the indicated half-cycle of the supply voltage, transformer T_1 is gating and transformer T_2 is resetting. The average reset voltage $\bar{e}_r'$ applied to the control winding of T_2 is the sum of three voltages. These are the signal voltage $\bar{e}_c$, the intrinsic delayed positive-feedback voltage $\bar{e}_g'$ produced by the series-connected control windings, and the intrinsic negative-feedback voltage $\bar{i}_cR_c$. This last voltage is due to the summation of the IR drops produced by core excitation current (dictated by the excitation impedance $\bar{R}$) and, in the simplified block diagram (Fig. 2), the transformer-coupled load current $\bar{i}_L$. The reset voltage referred to the load circuit $\bar{e}_r$ in the given half-cycle is equal to the gate voltage $\bar{e}_g$ effective in the subsequent half-cycle; thus the half-cycle time delay ϵ^{-Ts} is introduced using Laplace notation. The average output voltage $\bar{e}_o$ is the difference between the average supply voltage E_s and the sum of the gate voltage $\bar{e}_g$ and the reset voltage $\bar{e}_r$. The load current $\bar{i}_L$ is the ratio of output voltage to the total resistance in the load circuit (R_L+r_x), and the load voltage, of course, is the product of the load current $\bar{i}_L$ and the load resistance R_L.

A steady state is achieved when the gating voltage in one half-cycle equals the gating voltage in the subsequent half-cycle. Therefore, in a given half-cycle under steady-state conditions, the gating voltage $\bar{e}_g'$ must equal the reset voltage

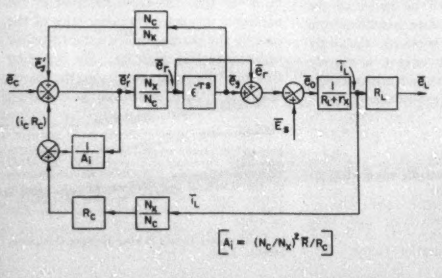

$\bar{e}_c$
$\bar{e}_g'$
$(i_c R_c)$
$\bar{e}_r'$
N_x/N_c
N_c/N_x
ϵ^{-Ts}
$\bar{e}_r$
$\bar{e}_g$
$\bar{e}_o$
E_s
$\frac{1}{R_L+r_x}$
$\bar{i}_L$
R_L
$\bar{e}_L$
$\frac{1}{A_i}$
R_c
$[A_i = (N_c/N_x)^2 \bar{R}/R_c]$

Fig. 2. Transformed block diagram for the series saturable reactor

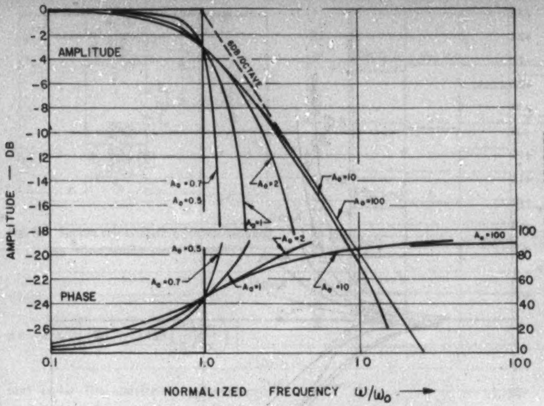


Fig. 4. Phase-amplitude plot of G(jω) for the saturable reactor

$\bar{e}_r'$. From the block diagram it can be seen that a necessary condition for a steady state is that the sum of the signal voltage $\bar{e}_c$ and the intrinsic negative-feedback voltage $\bar{i}_c R_c$ be zero. This is, of course, exactly what one expects. The symmetrical nature of the circuit makes the block diagram valid for both half-cycles of the supply voltage.

Observations Based on Block Diagram

Although the block diagram of Fig. 2 affords a simpler introduction to this technique, the block diagram of Fig. 1(B) is the more generally useful form. Thus, the average reset voltage $\bar{e}_r'$ across the control winding is determined by the ratio of the impedance presented by that winding to the external resistance in the control circuit R_c. In the self-saturating circuit (Fig. 8) this impedance ratio is simply the ratio of reflected excitation impedance to control-circuit resistance, i.e., $A_i = \hat{R}(N_c/N_x)^2/R_c$. However, in the case of the saturable reactor, the decoupling action provided by self-saturating rectifiers is absent. As a result, the impedance presented by the control winding is the parallel combination of the reflected excitation impedance $\hat{R}(N_c/N_x)^2$ and the reflected load-circuit resistance $(R_L+r_x)(N_c/N_x)^2$, modified by the delayed negative-feedback effect of the gate voltage (via the lower feedback loop). Unlike the self-saturating circuit, two intrinsic amplification factors must be defined; namely, $A_i = \hat{R}(N_c/N_x)^2/R_c$ and $A_i' = (R_L+r_x)(N_c/N_x)^2/R_c$.

It is of interest to note that the time constant normally associated with saturable reactors does not appear in the proposed block diagrams. It will be shown that this familiar lag is a direct result of the intrinsic positive-voltage feedback produced by the series-connected control windings. This particular feedback loop is shown on the block diagram as the gate voltage $\bar{e}_g$ fed back through the turns ratio N_c/N_x into the control-circuit summing point.

One interesting distinction between the self-saturating class and the saturable-reactor class of magnetic amplifiers is clearly illustrated by the block diagram. In the case of the latter class, there is a forward transmission path which bypasses the delay ϵ^{-T_s} and contributes directly to the output. Thus, although a minimum response time of one half-cycle is inevitable in the case of the self-saturating amplifier, there is no such limit on the speed of response of the saturable reactor. The common assumption that the response of the saturable reactor to a step-function input may be represented, for relatively slow response amplifiers, as a pure delay and single time constant will be shown to be incompatible with the transfer function which derives from the block diagram. It will be seen subsequently that, as a result of the foregoing distinction, the gain-bandwidth product of the series saturable reactor is twice that of the self-saturating magnetic amplifier.

Parallel Saturable Reactor

The techniques developed for the series saturable reactor may also be applied to the parallel saturable reactor of Fig. 3(A) to yield the block diagram of Fig. 3(B). The similarity between the two types is more clearly illustrated by a comparison of their block diagrams than by a comparison of their circuit diagrams.

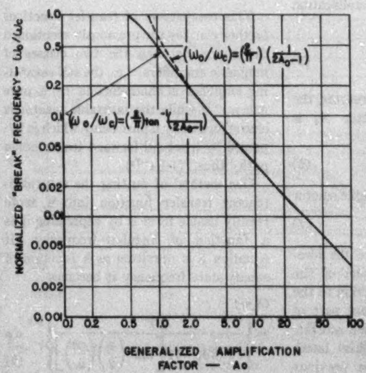


Fig. 5 (left). Normalized break frequency versus the generalized amplification factor A_0

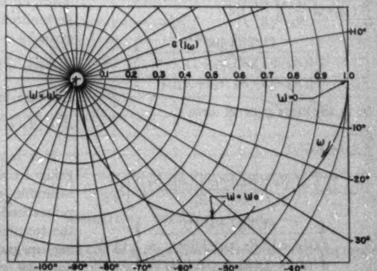


Fig. 6 (right). Polar co-ordinate plot of G(jω) for the saturable reactor

Most of the observations based on the block diagram of the series saturable reactor also apply to the parallel saturable reactor. However, there is one interesting distinction. Note that the delayed feedback introduced by the lower feedback loop of Fig. 3(B) is positive feedback, while in the series circuit the corresponding loop introduced negative feedback. It will be seen that this positive feedback raises the gain above what would normally be a very low value due to the fact that the impedance of the resetting control winding is shunted by the normally small reflected winding resistance r_x. This positive-feedback loop also serves to make the response time of the parallel saturable reactor considerably greater than that of the equivalent series saturable reactor.

Saturable-Reactor Transfer Functions

In a properly designed saturable reactor, the excitation impedance of the core is negligibly high compared with the load resistance, therefore, the intrinsic feedback loop involving this factor is omitted, i.e., $A_f \to \infty$, in the derivation of the transfer functions from the block diagrams.

The transfer function for the series saturable reactor may be written directly from the block diagram, for incremental quantities, as

$$\frac{\bar{e}_L}{\bar{e}_c}(s)=\left(\frac{N_c}{N_x}\right)\left(\frac{R_L}{R_c}\right)\left|\frac{1+\epsilon^{-Ts}}{(A_t'+1)-(A_t'-1)\epsilon^{-Ts}}\right| \qquad (1)$$

or in the more general form,

$$\frac{\bar{e}_o}{\bar{e}_c}(s)=KG(s)=\left[\left(\frac{N_x}{N_c}\right)(A_t')\right]\times \left[\frac{1+\epsilon^{-Ts}}{(A_t'+1)-(A_t'-1)\epsilon^{-Ts}}\right] \qquad (2)$$

In like manner, the transfer function for the parallel saturable reactor may be written from the block diagram of Fig. 3(B) as

$$\frac{\bar{e}_L}{\bar{e}_c}(s)=2\left(\frac{N_c}{N_x}\right)\left(\frac{R_L}{R_c}\right)\times \left\{\frac{(1/2)(1+\epsilon^{-Ts})}{[(A_t'+1)(R_L/r_x+1)]-[(A_t'+1)(R_L/r_x)+A_t']\epsilon^{-Ts}}\right\} \qquad (3)$$

or, in the more general form

$$\frac{\bar{e}_o}{\bar{e}_c}(s)=KG(s) =\left[2\left(\frac{N_x}{N_c}\right)\frac{(R_L+r_x/2)(N_c/N_x)^2}{R_c}\right]\times \left\{\frac{(1/2)(1+\epsilon^{-Ts})}{(A_t'+1)(R_L/r_x+1)]-[(A_t'+1)(R_L/r_x)+A_t']\epsilon^{-Ts}}\right\} \qquad (4)$$

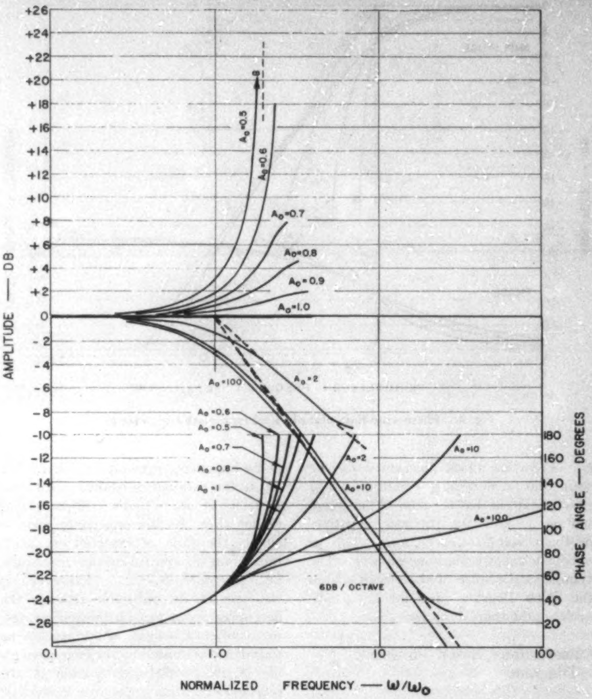


Fig. 7. Phase-amplitude plot of G(jω) for the self-saturating magnetic amplifier

It has been found that, in the interest of simplicity, the frequency-dependent part of the foregoing transfer functions can be reduced to the following general form, where A_o is the generalized amplification factor.

$$G(s)=\frac{1}{2}\left[\frac{(1+\epsilon^{-Ts})}{A_o-(A_o-1)\epsilon^{-Ts}}\right] \qquad (5)$$

Thus, for the series saturable reactor, the generalized amplification factor A_o is given by

$$A_o=(A_t'+1)/2 \qquad (6)$$

Likewise, for the parallel saturable reactor

$$A_o=(A_t'+1)(R_L/r_x+1) \qquad (7)$$

It is interesting to compare the foregoing frequency-dependent part of the saturable-reactor transfer function to the generalized frequency-dependent part of the transfer function for the self-saturating magnetic amplifier. This latter quantity, as taken from the previous paper,[8] is as follows:

$$G(s)=\left[\frac{\epsilon^{-Ts}}{(A_o)-(A_o-1)\epsilon^{-Ts}}\right],\quad A_o=(A_t+1) \qquad (8)$$

This comparison of transfer functions further clarifies the previously mentioned distinction between the two classes of magnetic amplifiers; i.e., the self-saturating amplifier is characterized by a pure delay ϵ^{-Ts} while the saturable reactor is characterized by a pure delay which is by-passed by a second forward transmission path, thus, $(1+\epsilon^{-Ts})$.

One method of putting the saturable-reactor transfer function into a more readily usable form is by expressing it as a function of angular frequency. If equation 5 is rewritten as a function of steady-state frequency it becomes

$$G(j\omega) =\left\{\frac{1}{1+(2A_o-1)^2\left[\tan^2\left(\frac{\pi}{2}\right)\left(\frac{\omega}{\omega_c}\right)\right]}\right\}^{1/2} <\phi \qquad (9)$$

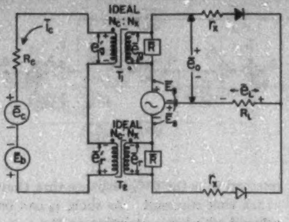


Fig. 8. Simple self-saturating (center-tapped) magnetic-amplifier circuit diagram

where

$$\phi = -\tan^{-1}\left[(2A_o-1)\tan\left(\frac{\pi}{2}\right)\left(\frac{\omega}{\omega_c}\right)\right] \tag{10}$$

and ω_c is the carrier (supply-voltage) angular frequency.

If ω_o is the frequency at which $\phi=45$ degrees, then

$$\frac{\omega_o}{\omega_c} = \left(\frac{2}{\pi}\right)\tan^{-1}\left(\frac{1}{2A_o-1}\right) \tag{11}$$

The substitution of equation 11 into equations 9 and 10 yields

$$G(j\omega) = \left\{\frac{1}{1+(2A_o-1)^2\tan^2\left[\left(\frac{\omega}{\omega_o}\right)\tan^{-1}\frac{1}{2A_o-1}\right]}\right\}^{1/2} < \phi \tag{12}$$

and

$$\phi = -\tan^{-1}(2A_o-1)\times \left\{\tan\left[\left(\frac{\omega}{\omega_o}\right)\tan^{-1}\frac{1}{2A_o-1}\right]\right\} \tag{13}$$

Equations 12 and 13 are plotted in Fig. 4, and equation 11 is plotted in Fig. 5. The phase-amplitude plot of Fig. 4 reveals some rather remarkable things about the saturable reactor. Clearly, in one limit ($A_o \rightarrow \infty$), the operation of the saturable reactor approaches that of a simple first-order system which is characterized by a single time constant. However, in the other limit ($A_o \rightarrow 0.5$), the operation is that of an ideal low-pass filter. The polar co-ordinate plot of $G(j\omega)$ which appears in Fig. 6 reveals the interesting fact that the relationship between amplitude and phase is that of a first-order system regardless of the value of A_o. Thus the attenuation is always 3 decibels when ϕ is 45 degrees, and the attenuation always approaches infinity as ϕ approaches 90 degrees.

A comparison of the foregoing phase-amplitude plot to that of the self-saturating magnetic amplifier, shown in Fig. 7 reveals the appreciable basic difference between the transfer functions of the two classes of magnetic amplifier. Obviously in one limit, i.e., $A_o \rightarrow \infty$, the operation of both devices approaches that of a simple first-order system which is characterized by a single time constant. However, as the operation departs from this limit, the self-saturating circuit exhibits less attenuation and more phase shift, while the saturable reactor is characterized, by both increased attenuation and increased phase shift (at frequencies higher than the break frequency ω_o). In the simple self-saturating circuit with no external feedback, shown in Fig. 8, as the circuit is forced toward fast re-

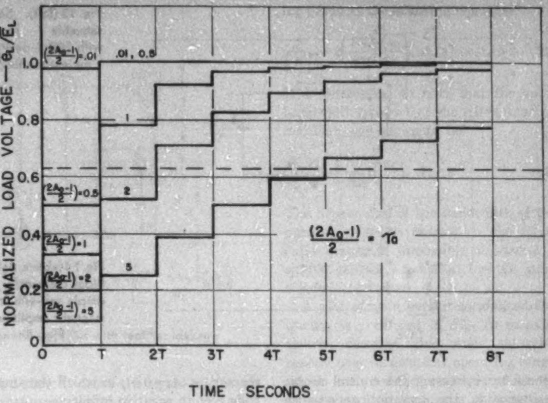


Fig. 9. Normalized response of the saturable reactor to a step-function driving voltage

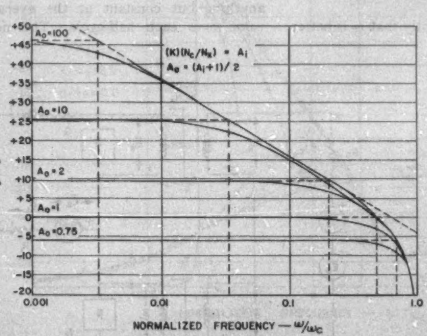


Fig. 10. Voltage gain versus normalized frequency for the series saturable reactor

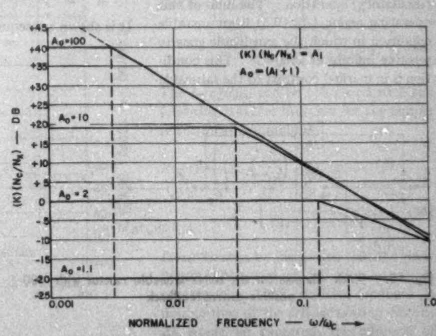


Fig. 11. Asymptotic plot of voltage gain versus normalized frequency for the simple self-saturating magnetic amplifier of Fig. 8

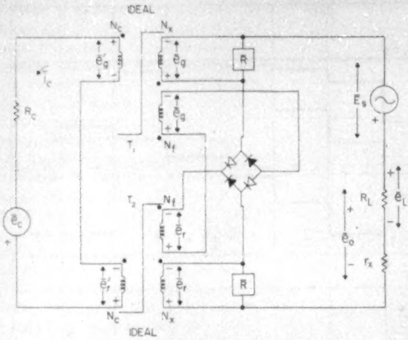

IDEAL
N_c
N_x
$\bar{e}_g$
R
$\bar{i}_c$
R_c
E_s
T_1
N_f
T_2
$\bar{e}_r$
$\bar{e}_r'$
$\bar{e}_c$
R_L
$\bar{e}_L$
$\bar{e}_o$
r_x
IDEAL

Fig. 12 (left). Series saturable reactor with 100% positive-current feedback

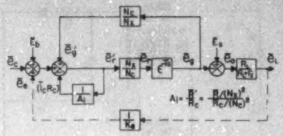

$\frac{N_c}{N_x}$
E_b
$\bar{e}_g'$
$\bar{e}_c$
$\bar{e}_e$
$(\bar{I}_cR_c)$
$\frac{1}{A_i}$
$\bar{e}_r'$
$\frac{N_x}{N_c}$
$\bar{e}_r$
ϵ^{-Ts}
$\bar{e}_s$
E_s
$\bar{e}_o$
$\frac{R_L}{R_L+r_x}$
$\bar{e}_L$
$A_i = \frac{R'}{R_c} = \frac{R/(N_x)^2}{R_c/(N_c)^2}$
$\frac{1}{K_s}$

Fig. 14 (above, right). Block diagram for the simple self-saturating magnetic amplifier of Fig. 8

sponse by increasing the control-circuit resistance R_c, the operation approaches the limit in which the intrinsic amplification factor A_i approaches zero, and the generalized amplification factor A_o becomes equal to one. When A_o is equal to one, the operation corresponds to a pure delay, i.e., phase shift with no attenuation. In the case of the series saturable reactor the analogous situation causes the generalized amplification factor A_o to approach one half. Therefore it can be seen that, in the case of the latter circuit, the limiting operation for fast response corresponds to an ideal low-pass filter. Clearly, if the common assumption that the self-saturating circuit is characterized by a pure time delay and a single time constant is valid, then this same assumption can not properly be applied to the saturable reactor.

It is interesting to note that the application of external feedback to the self-saturating amplifier can cause A_o to be less than one and produce underdamped (oscillatory) operation. The limit of this operating region ($A_o=0.5$) is an unstable condition in which the amplitude goes to positive infinite at $\omega/\omega_o=2$. This condition is in marked contrast to the saturable reactor (at $A_o=0.5$), in which the amplitude goes to negative infinity.

Saturable-Reactor Response Time

If $A_o \geq 5$, the relationship between break frequency ω_o and the generalized amplification factor A_o can be expressed with good approximation (see Fig. 5) as

$$\frac{\omega_o}{\omega_c}=\left(\frac{2}{\pi}\right)\left(\frac{1}{2A_o-1}\right),\ A_o \geq 5 \tag{14}$$

A study of Fig. 4 reveals that a first-order approximation is valid for values of A_o greater than ten. Therefore, an equivalent time constant for the saturable-reactor response to a step function can be written as follows

$$\tau_a=\frac{1}{\omega_o}=\left(\frac{\pi}{\omega_c}\right)\left(\frac{2A_o-1}{2}\right),\ A_o \geq 10 \tag{15}$$

Further, if one lets $\pi/\omega_c=T$, the time for a half-cycle of the carrier voltage, then

$$\tau_a=\left(\frac{2A_o-1}{2}\right)(T),\ A_o \geq 10 \tag{16}$$

It is shown subsequently that τ_a is better referred to as the 63% response time than as the time constant. As such, τ_a can be utilized with less restriction on the value of A_o.

The expression for response time can be put into terms of the individual saturable-reactor parameters by utilizing equations 6 and 7. Thus, for the series saturable reactor

$$\tau_a=(A_i'/2)T=\frac{1}{2}\left(\frac{R_L+r_x}{R_c}\right)\left(\frac{N_c}{N_x}\right)^2\left(\frac{\pi}{\omega_c}\right),\quad A_o \geq 10 \tag{17}$$

or

$$\tau_a=\frac{1}{4f}\left(\frac{R_L+r_x}{R_c}\right)\left(\frac{N_c}{N_x}\right)^2,\ A_o \geq 10 \tag{18}$$

For the parallel saturable reactor

$$\tau_a=\frac{1}{2}\left[2(A_i'+1)\left(\frac{R_L}{r_x}+1\right)-1\right]T,\ A_o \geq 10 \tag{19}$$

or

$$\tau_a=\frac{1}{4f}\left[4\left(\frac{R_L+r_x/2}{R_c}\right)\left(\frac{N_c}{N_x}\right)^2+\left(\frac{R_L+r_x/2}{r_x/2}\right)\right],\ A_o \geq 10 \tag{20}$$

The foregoing expressions are in agreement with the previous work of Storm.[6]

Step-Function Response

A consistent half-cycle average type of analysis precludes the representation of the saturable-reactor input or output as anything but constant at the average value over each half-cycle. The nor-

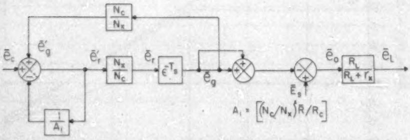

$\frac{N_c}{N_x}$
$\bar{e}_c$
$\bar{e}_g'$
$\bar{e}_r'$
$\frac{N_x}{N_c}$
$\bar{e}_r$
ϵ^{-Ts}
$\bar{e}_g$
$\bar{e}_o$
$\frac{R_L}{R_L+r_x}$
$\bar{e}_L$
$\bar{E}_s$
$\frac{1}{A_i}$
$A_i = [(N_c/N_x)^2 R/R_c]$

Fig. 13. Block diagram for the series saturable reactor with 100% positive-current feedback

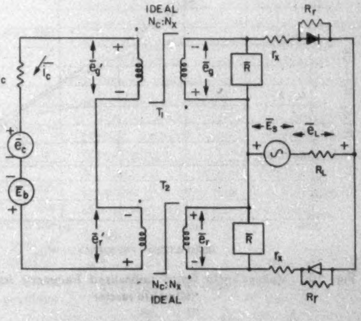

IDEAL
$N_c:N_x$
R_f
R_c
$\bar{i}_c$
$\bar{e}_g'$
$\bar{e}_g$
R
r_x
T_1
$\bar{E}_s$
$\bar{e}_L$
$\bar{e}_c$
R_L
$\bar{E}_b$
T_2
$\bar{e}_r'$
$\bar{e}_r$
$N_c:N_x$
IDEAL

Fig. 15 (right). Circuit diagram for self-saturating (doubler) magnetic amplifier with rectifier leakage

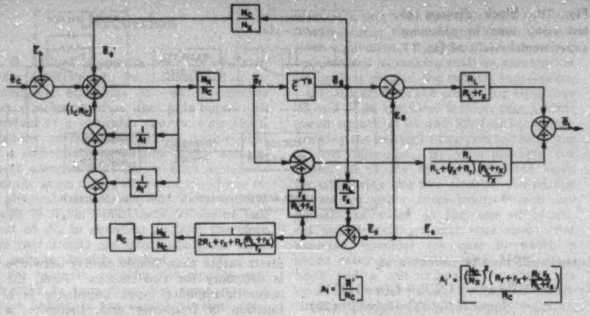


Fig. 16. Block diagram for self-saturating magnetic amplifier with rectifier leakage

malized response of the saturable reactor to a step-function input is presented in Fig. 9. It is of particular interest to note that for values of $(2A_o-1)/2$, i.e., τ_a/T, equal to or greater than one, i.e., $A_o \geq 1.5$, at least 63% response is achieved at time $t=[(2A_o-1)/2](T)$. Thus, τ_a can be referred to as the 63% response time for values of A_o equal to or greater than 1.5. Clearly there is no half-cycle delay associated with the response of the saturable reactor such as is characteristic of the self-saturating amplifier. In fact, the only lower limit on the response time of the series saturable reactor is a practical limit imposed by the maximum control-circuit resistance which can be utilized. Obviously, it must be realized that any change in output power which is achieved during the first half-cycle must be supplied directly from the signal source via transformer action.

Gain-Bandwidth Product

An important parameter associated with amplifiers in general is the gain-bandwidth product. If one defines the saturable-reactor bandwidth as corresponding to the break frequency ω_o, simple expressions for the gain-bandwidth product $K\omega_o$ can be derived. For the series saturable reactor

$$K=(N_x/N_c)(A_i)$$

and

$$\omega_o=\left(\frac{\omega_c}{\pi}\right)\left(\frac{2}{2A_o-1}\right),\ A_o\geq 5$$

Substituting $A_o=(A_i+1)/2$ and simplifying produces the following equation:

$$(K\omega_o)=2\left(\frac{N_x}{N_c}\right)\left(\frac{\omega_c}{\pi}\right),\ A_o\geq 5 \text{ or } A_i'\geq 9 \quad (21)$$

It is interesting to note that the gain-bandwidth product of the self-saturating amplifier was shown to be[3]

$$(K\omega_o)=\left(\frac{N_x}{N_c}\right)\left(\frac{\omega_c}{\pi}\right),\ A_i\geq 9 \quad (22)$$

The dependence of the bandwidth of the parallel saturable reactor on the R_L/r_x ratio makes it impossible to derive a simple general expression for the gain-bandwidth product.

A plot of gain versus normalized frequency is presented in Fig. 10 to more clearly describe the series saturable reactor over its complete operating range, which includes that region in which the gain-bandwidth product is not constant. For comparison purposes an asymptotic plot of the corresponding information which has been obtained from the simple self-saturating magnetic amplifier is presented in Fig. 11.

Series Saturable Reactor with External Feedback

The application of 100% positive-current feedback to the series saturable reactor is illustrated in Fig. 12 for the simple case in which the number of turns in the positive-feedback winding N_f is equal to the number of turns in the load winding N_x.

The block diagram for this amplifier, shown in Fig. 13, is almost identical to that of the self-saturating magnetic amplifier, shown in Fig. 14. The transfer function is

$$KG(s)=\left[(2A_i)\left(\frac{N_x}{N_c}\right)\right]\times \left[\frac{\epsilon^{-Ts}}{(A_i+1)-(A_i)\epsilon^{-Ts}}\right] \quad (23)$$

The foregoing can be compared with the following transfer function for the simple self-saturating amplifier.

$$KG(s)=\left[(A_i)\left(\frac{N_x}{N_c}\right)\right]\times \left[\frac{\epsilon^{-Ts}}{(A_i+1)-(A_i)\epsilon^{-Ts}}\right] \quad (24)$$

It can be seen by inspection that the gain-bandwidth product of the saturable reactor with 100% positive feedback is double that of the self-saturating circuit. As would be expected, the addition of feedback to the saturable reactor has no effect on its gain-bandwidth product.

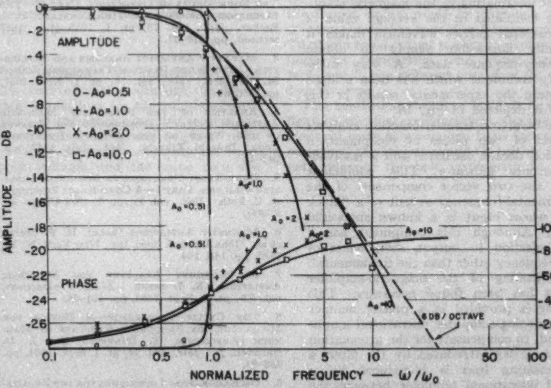


Fig. 17. Comparison of experimental operation with theoretical operation for the series saturable reactor

Parallel Saturable Reactor as a Self-Saturating Amplifier with Rectifier Leakage

An interesting comparison between the parallel saturable reactor and the self-saturating (doubler) magnetic amplifier can be obtained by treating the former circuit as the limiting case of a self-saturating circuit with rectifier leakage. The block diagram for the doubler circuit of Fig. 15 is presented in Fig. 16. Note that if the rectifier leakage resistance R_r goes to infinity, the block diagram reduces to that of the simple self-saturating circuit (Fig. 14). However, if the leakage resistance is zero, the block diagram becomes that of the parallel saturable reactor; see Fig. 3(B).

Experimental Verification of the Theory

The results of an effort to experimentally verify the validity of the transfer functions which have been derived are presented in Fig. 17. A brief discussion of the experimental techniques which have been used is presented in Appendix II.

Conclusions

A block-diagram method of analysis has been extended from the self-saturating class of magnetic amplifiers to the saturable-reactor class of magnetic amplifiers. It has been shown that a single theory of operation can be used to treat a variety of magnetic-amplifier circuits. The block-diagram method of analysis is submitted as a useful tool not only for the analysis of conventional circuits but as a means of synthesizing new types of circuitry.

Appendix I. Development of Block-Diagram for Series Saturable-Reactor

The voltage-loop equation for the control circuit, Fig. 1(A), during the $(n-1)$th half-cycle is

$$\bar{e}_g'(n-1)T+\bar{i}_c(n-1)TR_c-\bar{e}_c(n-1)T = \bar{e}_r'(n-1)T \quad (25)$$

In the load circuit, assuming $\bar{R} \gg (R_L+r_x)$,

$$\bar{e}_r(n-1)T+\bar{e}_g(n-1)T=E_s-\bar{e}_o(n-1)T \quad (26)$$

and

$$\bar{e}_r(nT)+\bar{e}_g(nT)=E_s-\bar{e}_o(nT) \quad (27)$$

Since

$$|\bar{e}_r'(n-1)T|=(N_c/N_x)|\bar{e}_r(n-1)T| =(N_c/N_x)|\bar{e}_g(nT)| \quad (28)$$

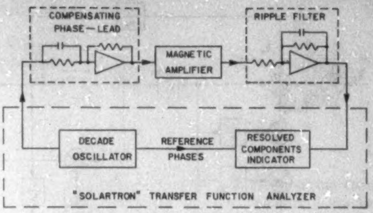

COMPENSATING PHASE — LEAD
MAGNETIC AMPLIFIER
RIPPLE FILTER
DECADE OSCILLATOR
REFERENCE PHASES
RESOLVED COMPONENTS INDICATOR
"SOLARTRON" TRANSFER FUNCTION ANALYZER

Fig. 18. Block diagram of test setup used in obtaining experimental results of Fig. 17

equation 25 may be rewritten as

$$[-e_c(n-1)T+\bar{i}_c(n-1)TR_c+\bar{e}_g(n-1)T]\times (N_x/N_c)=\bar{e}_g(nT) \quad (29)$$

With the use of equation 27 it can be seen that

$$\bar{e}_o(nT)=\bar{E}_s-\bar{e}_r(nT)-[-\bar{e}_c(n-1)T+ \bar{i}_c(n-1)TR_c+\bar{e}_g'(n-1)T](N_x/N_{cs}) \quad (30)$$

Taking the Laplace transform yields the following:

$$\bar{e}_o(s)=\bar{E}_s(s)-\bar{e}_r(s)-[-\bar{e}_c(s)+\bar{i}_c(s)R_c+ \bar{e}_g'(s)][N_x/N_c)\epsilon^{-Ts} \quad (31)$$

In Fig. 2, $\bar{i}_c(s)R_c=(N_x/N_c)\bar{i}_L(s)R_c$, etc.

Appendix II. Experimental Technique

The problem of experimentally determining the dynamic characteristics of the magnetic amplifier is a relatively difficult one which has been approached in a variety of ways in the past.[7-9] Although the magnetic amplifier most readily lends itself to the measurement of transient response, frequency-response information is normally more valuable for purposes of control system design. The fact that the output information of the magnetic amplifier is contained in the average value of the distorted carrier waveform makes it normally impossible simply to obtain frequency-response data. A very satisfactory technique which has been utilized to obtain the experimental results in this paper is depicted in Fig. 18.

The Solartron transfer-function analyzer consists of two pieces of equipment, a precision decade oscillator and a resolved-components indicator. This equipment yields the two vector components of the fundamental-frequency output of a "black box" whose input is a known sinusoidal signal. Although this equipment has a high rejection to output components of any frequency other than the fundamental, some filtering of the magnetic-amplifier output has been found necessary. This filtering is provided in a precise manner by an analog-computer operational amplifier and, to compensate for the attenuation and phase lag introduced by the filter, a compensating lead is introduced by a second operational amplifier between the decade oscillator and the magnetic-amplifier input. The use of operational amplifier filters rather than simple passive networks is necessary for two reasons. First, the magnetic-amplifier input impedance is a function of frequency and, therefore, a passive lead network on the input would see a variable rather than a constant load. Obviously this would not allow for proper compensation of the lag introduced by the output filter. The use of operational amplifiers effectively circumvents this problem. Second, the effective signal-source impedance of the oscillator, lead-network combination is frequency-dependent and therefore does not permit the simple determination of the effective resistance in the control circuit R_c. The extremely low (approximately 1 ohm) output impedance of the operational amplifier permits the assumption of a zero-impedance signal source.

This frequency-response method of evaluating the dynamic performance of magnetic amplifiers has been found to be not only valuable, but essential, in measurements on fast-response saturable reactors.

References

1. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

2. THE SERIES MAGNETIC AMPLIFIER: PART I—THE FOUR MODES OF OPERATION; PART II—THE MECHANISMS OF GAIN AND TIME CONSTANT, R. C. Barker. *Ibid.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 819–31.

3. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, Robert M. Hubbard. *Ibid.*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

4. ANALYSIS OF THE FULL-WAVE MAGNETIC AMPLIFIER CIRCUITS CONSIDERING THE CHANGE OF THE WIDTH OF THE DYNAMIC HYSTERESIS LOOP, Tadashi Kikuchi. *Ibid.*, July 1957, pp. 241–49.

5. CORE-RESET FUNCTIONS IN MAGNETIC-AMPLIFIER ANALYSIS, PART I—A CORE-RESET FUNCTION, G. C. Feth. *Ibid.*, vol. 77, pt. I, 1958 (*Paper no. 57-95*).

6. MAGNETIC AMPLIFIERS (book), H. F. Storm, et al. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 148, 164.

7. A TRANSIENT ANALYZER FOR MAGNETIC AMPLIFIERS, E. J. Smith. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 461–65.

8. THE CYCLIC INTEGRATOR—A DEVICE FOR MEASURING THE FREQUENCY RESPONSE OF MAGNETIC AMPLIFIERS, T. Dunnegan, Jr., J. D. Harnden, Jr. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 358–65.

9. OSCILLOGRAPHIC TECHNIQUES FOR THE EVALUATION OF MAGNETIC AMPLIFIER RESPONSE, D. L. Critchlow. *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 607–10.

Discussion

A. G. Milnes (Carnegie Institute of Technology, Pittsburgh, Pa.): The author by comparing equations 23 and 24 reaches the conclusion that the gain-bandwidth product of a saturable reactor with 100% positive current feedback is double that of a self-saturating circuit. Such a statement assumes that N_x in both equations has the same value, whereas comparison on a physical basis shows that the numerical value of N_x in equation 23 should be half that of N_x in equation 24, and therefore the two circuits are substantially the same in gain-bandwidth product.

For example, consider the circuits constructed in turn from the same pair of cores (which is the logical physical comparison that should be made) and assume that each core has N turns capable of carrying load current and N_c turns rated for control current. Then, in the self-saturating connection, $N_x = N$; but with the same cores rearranged in the series circuit, with the same supply voltage, the gate turns per core become $1/2N(=N_x$ in equation 23) and the 100% feedback winding accounts for the remaining $1/2N$ turns per core. The conclusion that follows is that the two circuits are similar in voltage-gain-bandwidth product.

Since positive feedback does not alter the voltage-gain-bandwidth product of the series saturable-reactor circuit, it may be suggested that the feedback be omitted thus permitting N_x in equation 23 to be increased. This however considerably reduces the current gain of the series circuit and so lowers the power-gain-bandwidth product which is a matter of practical concern, and invalidates this method of comparison.

The relative merits of self-saturating and series reactor connections in static and transient response have been of interest to designers for many years and discussion can be found in the literature on this subject. In particular, the equivalence of gain-bandwidth products has been derived by relatively simple time-constant analysis in past papers.

R. M. Hubbard: Mr. Milnes is correct in his observation that although the fundamental expressions for the gain-bandwidth product of the series saturable reactor and that of the self-saturating amplifier differ by a factor of two, the assumption of equal supply voltages and the same core size dictates a factor of two difference in load-winding turns, thus making the two expressions equivalent. However, since I do not feel justified in making such an assumption an implicit part of the subject expressions, I actually erred in not comparing equations 23 and 24 in the same fashion that I compared equations 21 and 22, i.e., by simply drawing the reader's attention to the factor of two difference. By thus allowing the reader to supply his own assumed basis for comparing the two circuits, the author allows for other assumptions which are equally as valid as the one of identical supply voltages and core sizes. For example, consider the case in which a series saturable reactor with 100% feedback and a self-saturating amplifier are operating under thermally limited conditions with equal sinusoidal supply voltages. At maximum output, the load-winding-current form factor of the former circuit will be 1.11, while that of the latter circuit will be 1.57. Thus the load winding of the self-saturating circuit must utilize wire of approximately 40% greater cross section than in the case of the series saturable reactor. It can be seen that if the load winding and feedback winding of the saturable reactor occupy an area A, then the load winding of the self-saturating circuit must occupy an area $1.4 \times A$. Thus, in this case, the assumption of identical core sizes is not necessarily justified and, in fact, to validate Mr. Milnes' argument one must select a core for the self-saturating circuit which has an equal cross-sectional area and 40% greater window area than that of the saturable reactor.

Again, consider the case in which the two circuits are operating under resistance-limited conditions, and one chooses to obtain maximum power output by substantially matching the load resistance to the load-winding resistance. Clearly, using identical cores and supply voltages, it can be seen that the total load-circuit resistance in the case of the saturable reactor with 100% feedback is double that of the self-saturating circuit. Thus, under matched conditions, the obtainable load power in the case of the former circuit would be only one half of that obtainable from the latter circuit. To obtain equal output powers as a common basis for comparison, one would again be forced to use different core sizes for the two circuits.

The conclusion therefore is that, in the light of practical considerations, the use of identical cores as a common basis for comparison of the subject circuits is neither the only one nor the best assumption to be made. Thus the assumption that the number of turns in the load winding of the series saturable reactor is one half that of the equivalent self-saturating circuit is by no means the rule, and one can not with assurance make a general statement about the ratio of the gain-bandwidth products of the two classes of circuits.

I take exception to Mr. Milnes' contention that the gain-bandwidth product is invalid as a means of comparing the saturable reactor with feedback to the one without it. I would heartily agree if he had said that this is not the only parameter which is valuable for such a comparison. The importance of figure of merit or dynamic power gain is a well-known fact and I failed to explicitly mention this parameter because, as it applies to the self-saturating circuit, it was treated in some detail in an earlier paper[1] and also because the dynamic power gain for the series saturable reactor is very simply the product of current gain and gain-bandwidth product. Most certainly it does not follow that if the gain-bandwidth product is invariant with feedback then the feedback is extraneous.

The important point is that a saturable reactor, upon the application of feedback, is as well behaved as any other class of amplifier, be it vacuum-tube, transistor, etc., i.e.:

1. Feedback does not alter the gain-bandwidth product.
2. Voltage feedback does not alter the current gain and therefore does not alter the dynamic power gain.
3. Current feedback does alter the current gain and thereby alters the dynamic power gain.

In conclusion, I would like to point out that my purpose in presenting this block-diagram method of analysis was to show that there is a single method of analysis which can be applied to all classes of magnetic-amplifier circuits to yield results which are substantially in agreement with the results obtained by a variety of other methods both theoretical and emperical. In addition to the advantage of not requiring a different method of analysis for each class of circuit, this technique, I feel, sheds some light on the operation in the regions of fast response, which are excluded from most other techniques.

REFERENCE

1. See reference 3 of the paper.

The Effective Feedback Factor of Self-Balancing Magnetic Amplifiers

W. A. GEYGER
MEMBER AIEE

Synopsis: The effective feedback factor of a magnetic amplifier is a measure for the stability of operation (drift). In conventional forms of external-feedback or self-saturating circuits, the drift is linearly proportional to the actual value of ampere-turn gain. In these circuits, the drift can be reduced by reducing the extrinsic-feedback factor. However, because the ratio of gain reduction over drift reduction is constant, the gain will be reduced at the same rate.

In self-balancing magnetic amplifiers, a very significant reduction of drift is obtained by reducing the effective (total) feedback factor. This has been achieved by combination of the dynamic gain-increasing effect of magnetic positive feedback and the drift-reducing effect of electric negative feedback, resistance-coupled into the control-circuit loop.

This paper is concerned with the determination of the effective (total) feedback factor of self-balancing magnetic amplifiers. The basic considerations may be applied equally well to single-ended and push-pull circuits utilizing either external feedback or self-saturation. Typical examples of voltage-balance and current-balance forms of self-balancing magnetic-amplifier circuits are presented to demonstrate the practical significance of the simple equations with regard to design considerations on the stability of operation of such circuits.

THE ordinary saturable reactor is characterized by the fact that, over certain ranges, the average value of load current is substantially independent of large variations in magnitude, frequency, and waveshape of the power-supply voltage, and considerable changes in load resistance. With regard to these unique features, it is possible to employ ordinary saturable reactors with high-permeability rectangular-hysteresis-loop core material as d-c instrument transformers for measurement of large direct currents and high direct voltages.[1,2] This possibility is based upon the current-transformer character (law of equal ampere-turns[3]) which is the result of inherent negative-feedback properties (intrinsic negative feedback) of the ordinary saturable reactor having either series- or parallel-connected a-c load windings.

In conventional external-feedback or self-saturating types of magnetic-amplifier circuits, extrinsic positive feedback is provided for reducing or nearly nullifying the intrinsic negative feedback of the ordinary saturable reactor to obtain certain static and dynamic performance characteristics, as expressed, e.g., by ampere-turn gain, power gain, control time constant, and dynamic power gain. In these types of magnetic-amplifier circuits, the effective feedback ratio[4] or total feedback factor[3] is a measure for the stability of operation (drift). The drift is linearly proportional to the ampere-turn gain and may be reduced by reducing the extrinsic-feedback factor. However, because the ratio of gain reduction over drift reduction is constant, the gain will be reduced at the same rate.

The aim of the research and developmental work on magnetic amplifiers of the self-balancing potentiometer type[5,6] consisted in combining the inherently extremely stable operating conditions of the ordinary saturable reactor, as used in d-c instrument transformers,[1,2] with the unique features of high-precision self-balancing potentiometer devices and negative-feedback amplifiers. In these self-balancing magnetic amplifiers, extrinsic magnetic positive feedback (external feedback or self-saturation) is provided for completely nullifying the intrinsic negative feedback of the ordinary saturable reactor (infinite internal gain or critical regeneration[5,6]). Additionally, extrinsic electric negative feedback (resistance-coupled into the control-circuit loop) is applied to obtain very stable operating conditions of the amplifier, similar to those of d-c instrument transformers. In self-balancing magnetic amplifiers, a very significant reduction of drift is obtained by reducing the effective (total) feedback factor. This has been achieved by combination of the dynamic gain-increasing effect of magnetic positive feedback and the drift-reducing effect of electric negative feedback. Operating under voltage-balance or current-balance conditions of the control circuit, these amplifiers demand practically no energy from the d-c signal source. Because the average value of the control-winding current is substantially zero for the entire working range of the amplifier, the linearity error is in the order of 1/10 of 1% or less. Over certain ranges, large variations in magnitude and frequency of power-supply voltage, load resistance, and control-circuit resistance have no appreciable effect upon the transconductance, voltage ratio, current ratio, or transresistance. Such amplifiers operate under very stable conditions with a large over-all power gain and a high speed of response. When using 100% negative feedback,[7,8] actual response time may be reduced to its minimum value of one half-cycle of the power-supply frequency. On the other hand, for special applications, the transient-response characteristics may be altered materially through the use of differentiating or integrating circuit elements in the degeneration network of the amplifier. Results of various experimental investigations show that the actual behavior of self-balancing magnetic amplifiers approximates very closely corresponding characteristics of an ordinary saturable reactor with stabilizing negative-feedback properties (intrinsic negative feedback).

This paper is concerned with the determination of the effective (total) feedback factor of self-balancing magnetic amplifiers. The basic considerations may be applied equally well to single-ended and push-pull circuits utilizing either external feedback or self-saturation. Typical examples of voltage-balance and current-balance forms of self-balancing magnetic-amplifier circuits are presented to demonstrate the practical significance of the simple equations with regard to design considerations on the stability of operation of such circuits.

Effective Feedback Factor of Single-Ended Self-Balancing Magnetic Amplifiers

VOLTAGE-BALANCE EXTERNAL-FEEDBACK CIRCUIT

In the circuit of Fig. 1, the d-c signal voltage E_S is balanced by the compensating voltage $E_K = I_L R_K$. The average value of control-winding current is, if I_C is very small in comparison with I_L,

$$I_C=\frac{E_S-E_K}{R_C+R_K}=\frac{E_S}{R_C+R_K}-\frac{E_K}{R_C+R_K}=I_S-I_K \quad (1)$$

where E_S is the electromotive force of

Paper 57-767, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 11, 1957; made available for printing July 9, 1957.

W. A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

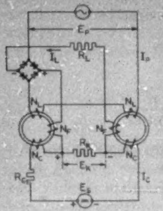


Fig. 1. Self-balancing magnetic amplifier of voltage-balance external-feedback type. Transconductance of this circuit is $I_L/E_S = 1/R_K$, and effective feedback factor is $h = -(N_C/N_L)R_K/(R_C+R_K)$

the d-c signal source, and R_C is the total resistance of the control circuit between the terminals of negative-feedback resistor R_K (including the sum of the resistances of the control-winding units N_C, and the internal resistance of the signal source). The term $I_S = E_S/(R_C+R_K)$ is the signal-current component, and $I_K = E_K/(R_C+R_K)$ is the compensating-current component.

It was shown[5] that the relationship between average value of load current I_L and signal voltage E_S is given by the equations

$$\frac{E_S - I_L R_K}{R_C+R_K} = I_L \frac{N_L - N_F}{N_C} \tag{2}$$

or

$$\frac{E_S}{I_L} = R_K + (R_C+R_K)\frac{N_L-N_F}{N_C} \tag{3}$$

Thus, with $N_F = N_L$, the following results

$$\frac{E_S}{I_L} = R_K \text{ or } \frac{I_L}{E_S} = \frac{1}{R_K} \tag{4}$$

where $1/R_K$ represents the transconductance of the circuit. In this case

$$I_C = I_S - I_K = 0 \tag{5}$$

The condition, $N_F = N_L$, corresponding to critical regeneration,[8] involves the assumption of ideal properties of the saturable-reactor elements and feedback rectifiers. This assumption means that the rectifier output current I_L has exactly the average value of the rectifier input current I_P. In practice, however, departure of the saturable-reactor and rectifier components from the ideal characteristics may cause a certain reduction of current ratio I_L/I_P so that actual external-feedback ampere-turns $I_L N_F$ are correspondingly reduced.

To compensate for imperfections of the saturable reactors and rectifiers, the number of turns of external-feedback windings N_F may be correspondingly increased. Another practical possibility consists in providing an additional positive-feedback circuit containing a current-divider network with boosting windings. In this way, it is possible to apply fractional turns[9] for proper adjustment ($I_C = 0$).

With regard to the following considerations, the method of obtaining the desired critical-regeneration condition is not important. Actually, in any case, the extrinsic positive-feedback ampere-turns will be adjusted so that the intrinsic negative feedback of the saturable reactor is completely nullified, and the average value of I_C is substantially zero. Therefore, to simplify the following equations, it is assumed that critical regeneration ($I_C = 0$) is obtained with $N_F = N_L$.

The magnitude of the intrinsic negative feedback of the ordinary saturable reactor can be expressed by the intrinsic-feedback factor $h_i = -1$, as defined by H. F. Storm,[3] where h_e designates the extrinsic-feedback factor, and $h = h_e + h_i = h_e - 1$ represents the total or effective feedback factor of the saturable reactor with extrinsic positive feedback. The special form of compound-feedback circuitry, as used in the circuit of Fig. 1, utilizes two separate and fundamentally different extrinsic-feedback components, and therefore involves two separate extrinsic-feedback factors: the positive extrinsic-feedback factor $h_{e,m}$ corresponding to the positive magnetic feedback, which is produced by the external-feedback windings N_F; and the negative extrinsic-feedback factor $h_{e,e}$ corresponding to the negative electric feedback, which is produced by the control windings N_C acting as negative-feedback windings.

In accordance with Storm's concept,[3] the effective or total feedback factor of the circuit, Fig. 1, is defined as follows:

$$h_i = -1,\ h_{e,m} = I_L N_F / I_P N_L = N_F/N_L = +1,$$

and

$$h_{e,e} = -I_K N_C / I_P N_L = -(I_L N_C / I_P N_L) \times R_K/(R_C+R_K) = -(N_C/N_L)R_K/(R_C+R_K)$$

Therefore

$$h = h_i + h_{e,m} + h_{e,e} = h_{e,e} = -\frac{N_C}{N_L}\frac{R_K}{R_C+R_K} \tag{6}$$

Furthermore, the remaining amount of applied positive feedback may be characterized by the resultant extrinsic-feedback factor

$$h_e = h_{e,m} + h_{e,e} = 1 - \frac{N_C}{N_L}\frac{R_K}{R_C+R_K} \tag{7}$$

When expressing the remaining amount of applied positive feedback in per cent,[10] the following equation may be used:

$$F_{eff} = +100\left(1 - \frac{N_C}{N_L}\frac{R_K}{R_C+R_K}\right)\% \tag{8}$$

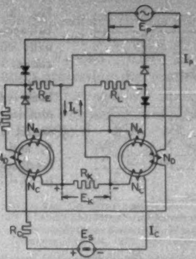


Fig. 2. Self-balancing magnetic amplifier of voltage-balance self-saturating type. Transconductance of this circuit is $I_L/E_S = 1/R_K$, and effective feedback factor is $h = -2(N_C/N_A)R_K/(R_C+R_K)$

The practical significance of these simple relationships with regard to design considerations may be demonstrated by three typical examples:

1. With $(N_C/N_L)R_K/(R_C+R_K) = 1.0$, the effective feedback factor is $h = -1$, and $F_{eff} = 0$. In this case, with regard to stability of operation, the behavior of the circuit of Fig. 1 is equivalent to that of the ordinary saturable reactor which is characterized by the intrinsic-feedback factor $h = h_i = -1$, i.e., the circuit has maximum stability of operation.
2. With $(N_C/N_L)R_K/(R_C+R_K) = 0.1$, the effective feedback factor is $h = -0.1$, the resultant extrinsic-feedback factor is $h_e = +0.9$, and $F_{eff} = +90\%$ corresponding to the turns ratio $N_F/N_L = 0.9$ of a conventional-type external-feedback circuit.[10]
3. With $R_K = 0$ and $h = 0$, i.e., when the electric negative feedback is completely removed, the circuit of Fig. 1 becomes unstable ("trigger effect" or "snap action").

VOLTAGE-BALANCE SELF-SATURATING CIRCUIT

In the circuit of Fig. 2, magnetic positive feedback is produced by self-saturation.[11] Electric negative feedback is produced by the control windings N_C acting as negative-feedback windings. Additionally, a small amount of magnetic positive feedback is produced by the external-feedback circuit $R_E R_D N_D$ to compensate for imperfections of the saturable reactors and rectifiers. In this circuit, critical-regeneration condition ($I_C = 0$) is obtained by adding the positive extrinsic-feedback effect of the self-saturating circuit and the small additional positive-feedback effect of the "boosting circuit" $R_E R_D N_D$. In this case, the effective or total feedback factor is defined in a similar way:

$$h_i = -1,\ h_{e,m} = +1,$$

and

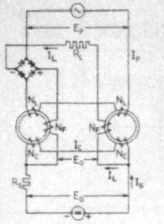


Fig. 3. Self-balancing magnetic amplifier of current-balance external-feedback type. Current ratio of this circuit is $I_L/I_S=1$, and effective feedback factor is $h=-N_C/N_L$

$$h=h_{e,e}=-2\frac{N_C}{N_A}\cdot\frac{R_K}{R_C+R_K} \tag{9}$$

The only difference is that in the circuit of Fig. 1 the total amount of magnetic positive feedback is produced by the external-feedback windings N_F (and sometimes by additional external-feedback windings providing the fractional turns[9]), whereas in the circuit of Fig. 2 the major part of magnetic positive feedback is produced by the self-saturating circuit, i.e., by means of the gate windings N_A with the associated half-wave rectifier elements. However, the corresponding positive extrinsic-feedback factor, $h_{e,m}=+1$, is the same in both cases.

CURRENT-BALANCE EXTERNAL-FEEDBACK CIRCUIT

In the circuit of Fig. 3, the signal current I_S is balanced by the opposing load current I_L, and the control-winding current $I_C=I_S-I_L$ is substantially reduced to zero as a result of critical regeneration ($N_F=N_L$). This arrangement operating under current-balance conditions of the control circuit is characterized by the fact that the effective input resistance of the magnetic amplifier is greatly reduced, and the actual voltage drop across the control windings N_C is substantially reduced to zero. In contrast to the voltage-balance circuits, Figs. 1 and 2, which act as voltmeters having very high resistances, the current-balance circuit of Fig. 3 acts as an ammeter having a very low resistance and a very small voltage drop across its terminals.

The effective feedback factor of the circuit in Fig. 3 is given by

$$h_i=-1,\ h_{e,m}=N_F/N_L=+1,$$

and

$$h_{e,e}-\frac{N_C}{N_L}$$

Therefore,

$$h=h_i+h_{e,m}+h_{e,e}=h_{e,e}=-\frac{N_C}{N_L} \tag{10}$$

With $N_C=N_F=N_L$, $h_{e,e}=-1$, and $F_{eff}=0$, i.e., the circuit has maximum stability of operation. It is to be noted that the total amount of load current I_L is superposed upon the signal current I_S. Thus, the current ratio $I_L/I_S=1$. The unique feature of this circuit is the fact that the input voltage E_0 of the amplifier is reduced to an extremely small value which is about 1,000 times smaller than the voltage drop across the control windings of the ordinary saturable reactor. However, the stability of operation, as characterized by the effective feedback factor $h=h_i=h_{e,e}=-1$, is the same in both cases.

CURRENT-BALANCE SELF-SATURATING CIRCUIT

In the circuit of Fig. 4, similar to that of Fig. 2, magnetic positive feedback is produced by self-saturation; and electric negative feedback is produced by the control windings N_C acting as negative-feedback windings. Additionally, a small amount of magnetic positive feedback is produced by the external-feedback circuit $R_ER_DN_D$ to compensate for imperfections of the saturable reactors and rectifiers. In contrast to the current-balance circuit of Fig. 3, a well-defined fraction of load current I_L is superposed upon the signal current I_S to increase the current ratio or current gain I_L/I_S of the amplifier:

$$\frac{I_L}{I_S}=\frac{R_1+R_2}{R_1}=\frac{I_L}{I_K} \tag{11}$$

This current ratio can be conveniently adjusted by a current-divider potentiometer resistor (Ayrton-Mather shunt). The control-winding current $I_C=I_S-I_K$ is substantially reduced to zero, and $E_0=0$.

The effective feedback factor of the circuit of Fig. 4 is given by

$$h_i=-1,\ h_{e,m}=+1,$$

and

$$h=h_{e,e}=-2\frac{N_C}{N_A}\,\frac{R_1}{R_1+R_2} \tag{12}$$

Therefore, e.g., with $2(N_C/N_A)R_1/(R_1+R_2)=0.05$, the effective feedback factor is $h=-0.05$, the resultant extrinsic-feedback factor is $h_e=+0.95$, and $F_{eff}=+95\%$, corresponding to the turns ratio $N_F/N_L=0.95$ of a conventional-type external-feedback circuit.[13]

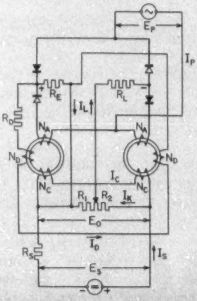


Fig. 4. Self-balancing magnetic amplifier of current-balance self-saturating type. Current ratio of this circuit is $I_L/I_S=(R_1+R_2)/R_1$, and effective feedback factor is $h=-2(N_C/N_A)R_1/(R_1+R_2)$

Effective Feedback Factor of Push-Pull-Type Self-Balancing Magnetic Amplifiers

GENERAL

The preceding considerations may be easily extended to the various forms of push-pull-type self-balancing magnetic amplifiers utilizing either external feedback or self-saturation for obtaining the critical-regeneration condition: $h_{e,m}=+1$. Generally, in such arrangements containing two single-ended external-feedback or self-saturating circuits in back-to-back connection, each of these circuits has its own positive-feedback component which completely nullifies the intrinsic negative feedback ($h_i=-1$) of the saturable reactor.[5,6,13–16] In another arrangement[17] with two separately biased saturable reactor systems utilizing differential feedback,[18,19] the resultant positive-feedback component of the magnetic amplifier is proportional to the actual difference between the individual load currents and is zero with no-signal conditions. This type of push-pull arrangement has proved to be very suitable for magnetic instrument amplifiers to be used in the realm of instrumentation, control engineering, and analog-computer techniques.

When designing such self-balancing magnetic amplifiers of the differential-feedback type,[17] determination of the effective feedback factor is of utmost importance. The larger the amount of negative-feedback ampere-turns (I_KN_C), the better is the stability of the amplifier with regard to zero drift and constant gain (linearity error). Therefore, such amplifiers are a representative example for illustrating the considerations on push-pull circuitry.

DETERMINATION

With respect to the determination of the effective feedback factor of differential-feedback self-balancing magnetic amplifiers, it was shown[17] that in such

Vol,
Pt. 1
1958

amplifiers critical regeneration is obtained with $N_F = N_L/2$. Thus, in this special case, the effective feedback factor is given by

$$h_i = -1,\ h_{e,m} = 2N_F/N_L = +1,$$

and

$$h_{e,e} = -2(N_C/N_L)R_K/(R_C+R_K)$$

Therefore

$$h = h_i + h_{e,m} + h_{e,e} = -2\,\frac{N_C}{N_L}\,\frac{R_K}{R_C+R_K} \qquad (13)$$

Evidently, the only difference between equations 6 and 13 is that in the single-ended external-feedback circuit of Fig. 1

$$h_{e,m} = N_F/N_L = +1,$$

and

$$h_{e,e} = -(N_C/N_L)R_K/(R_C+R_K),$$

whereas in the differential-feedback push-pull circuit,[17]

$$h_{e,m} = 2N_F/N_L = +1,$$

and

$$h_{e,e} = -2(N_C/N_L)R_K/(R_C+R_K)$$

Influence

The influence of the mixing network on the effective feedback factor of push-pull-type self-balancing magnetic amplifiers is discussed in the following. The relationship of equation 13 is based upon the assumption that a parallel mixing network or a twin-load mixing network is used for deriving the resultant polarity-reversible load current I_L from the individual load currents I_L' and I_L'' of the two saturable-reactor systems.[20] In this case:

$$I_L' = I_Q' + \Delta I_L',\quad I_L'' = I_Q'' - \Delta I_L'',$$

and

$$I_L = I_L' - I_L'' = 2\Delta I_L \qquad (14)$$

where $I_Q' = I_Q''$ are the quiescent-current values (the individual load currents with no-signal conditions) of the two saturable-reactor systems. Evidently, the factor 2 in equation 14 is responsible for the difference between equations 6 and 13.

When employing a series mixing network with mixing resistors R_S and load resistor R_L, the relationship between I_L', I_L'', and I_L is given by the equation[20]

$$I_L = (I_L' - I_L'')\,\frac{R_S}{2R_S+R_L} = 2I_L\,\frac{R_S}{2R_S+R_L} \qquad (15)$$

In this case, the negative-feedback ampere-turns (I_KN_C) are reduced by the resistance ratio $R_S/(2R_S+R_L)$ of the series mixing network with the passive resistance elements R_S; thus

$$h_{e,e} = -2\,\frac{R_S}{2R_S+R_L}\,\frac{N_C}{N_L}\,\frac{R_K}{R_C+R_K} \qquad (16)$$

These considerations illustrate the advantage of the use of parallel-type or twin-load mixing networks in push-pull-type self-balancing magnetic amplifiers.

Design Considerations

The relationship between the static characteristics (linearity error, power gain, zero-drift stability) and the transient performance (response time) of self-balancing magnetic amplifiers is such that if one property is accentuated, it is always at the expense of the other properties.[5] For example, it is generally desirable, in the interest of optimum linearity, to use as many control-winding turns (N_C) as the required response time will permit. In this way the inevitable residue of compensating ampere-turns which must be supplied by the control circuit will require a minimum of current (I_C) from that source, resulting in a minimum of linearity error.[18] Furthermore, the use of the greatest possible number of turns of N_C has the important added advantage of improving the stability of the amplifier by providing for a maximum of negative-feedback ampere-turns (I_KN_C).

On the other hand, in order to reduce actual response time of self-balancing magnetic amplifiers to its minimum value of one half-cycle of the power-supply frequency by introducing 100% negative feedback,[7,8] the number of turns of N_C must be reduced to smaller values. Analysis of the voltage-balance form of this amplifier[21] reveals that half-cycle response will be obtained if turns ratio $N_C/2N_L$ of external-feedback circuits (or the corresponding turns ratio N_C/N_A of self-saturating circuits) is approximately equal to the resistance ratio R_K/R_0 (the over-all voltage gain), where R_K is the actual value of the negative-feedback resistor, and R_0 is the total resistance of the output circuit including load resistance R_L, output-winding resistance, and forward resistance of the feedback rectifiers. Analysis of the current-balance form of the self-balancing magnetic amplifier reveals similar relationships.

Referring to practical values of the effective feedback factor of self-balancing magnetic amplifiers, it is to be noted that the resultant extrinsic-feedback factor, equation 7, characterizing the remaining amount of applied positive feedback, should not exceed values in the range $h_e = h_{e,m} + h_{e,e} = 0.95\ldots0.98$. This range of h_e corresponds to the turns ratio $N_F/N_L = 0.95\ldots0.98$ of a conventional external-feedback circuit ($F_{eff} = +95\%\ldots98\%$). With regard to this limitation, it is well to remember that with conventional external-feedback circuits utilizing a moderate amount of positive extrinsic feedback ($N_F/N_L = 0.5\ldots0.95$), the simple equation expressing the relationship between ampere-turn gain and turns ratio N_F/N_L holds to a good degree of accuracy (about 1% to 2%).[13]

References

1. Ein Einfacher Gleichstrom-Messwandler mit Echten Stromwandler-Eigenschaften, W. Krämer. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 58, 1937, pp. 1309–13.

2. Ein Neuer Gleichspannungs-Messwandler zur Messung Hoher Gleichspannungen, W. Krämer. *Ibid.*, vol. 59, 1938, pp. 1295–98.

3. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 77–79, 215–19.

4. The Effective Feedback Ratio of Magnetic Amplifiers, L. A. Finzi, G. F. Pittman, Jr., H. L. Durand. *AIEE Transactions*, vol. 71, pt. I, Apr. 1952, pp. 157–64.

5. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *Ibid.*, 1952 (Jan. 1953 section), pp. 383–95.

6. Self-Balancing Magnetic Amplifier, W. A. Geyger. *U. S. Patent no. 2,700,130*, Jan. 18, 1955.

7. On the Mechanics of Magnetic Amplifier Operation, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

8. Fast Response with Magnetic Amplifiers, D. G. Scorgie. *Ibid.*, vol. 72, pt. I, Jan. 1954, pp. 741–49.

9. 60-Cycle Self-Balancing Magnetic Servo Amplifier, W. A. Geyger. "Conference on Magnetic Amplifiers," *AIEE Special Publication T-86*, July 1956, pp. 221–34.

10. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 79–95.

11. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 835–50.

12. Magnetic Amplifiers with Orthonol Tape Cores, W. A. Geyger. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 59–67.

13. An Operational-Type Magnetic Amplifier for Air-Borne Servo-Control Systems, R. M. Hubbard. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 425–30.

14. Self-Balancing Magnetic Servo Amplifiers, W. A. Geyger. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 11, 1955, pp. 987–1000.

15. Self-Balancing Magnetic Servo Amplifier, W. A. Geyger. *Electronics*, New York, N. Y., vol. 29, no. 3, 1956, pp. 196–99.

16. Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load, W. A. Geyger. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 242–48.

17. Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type, W. A. Geyger. *Ibid.*, vol. 76, pt. I, Mar. 1957, pp. 39–46.

18. Einrichtung zum Messen von Kleinen Gleichstromgrössen, W. A. Geyger. *German Patent no. 702,082*, Jan. 30, 1941.

19. Ein Hochempfindlicher Magnetischer Nullstrom-Verstärker für die Mess-und Regeltechnik, W. A. Geyger. *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, Berlin-Siemensstadt, Germany, vol. 21, 1943, pp. 47–57.

20. Magnetic Amplifiers of the Balance Detector Type—Their Basic Principles, Characteristics, and Applications, W. A. Geyger. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1707–20.

21. The Operation of the Self-Balancing Magnetic Amplifier, Albert D. Krall, Edward T. Hooper. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

Shifting Counters

C. ELDERT
NONMEMBER AIEE

H. J. GRAY, JR.
NONMEMBER AIEE

H. M. GURK
NONMEMBER AIEE

MORRIS RUBINOFF
ASSOCIATE MEMBER AIEE

THIS PAPER DESCRIBES a problem and one solution to the problem which arose during the logical design of a digital computer for the real-time simulation of airplane flight. The computer in question is a 1-address machine with magnetic core memory which performs the basic operations of addition, comparison, transfer from arithmetic unit to memory, etc., at the rate of one operation every 5 microseconds. The logic of the machine is realized in pulse-type diode switching circuits similar to those employed by the National Bureau of Standards Eastern Automatic Computer and operating at an allowable pulse repetition frequency of 1.2 megacycles.

Because a 1-address mode of operation had been selected for the computer, it was necessary to provide a sequencing counter to specify the address of the next instruction upon the completion of the preceding order. Whether the sequence of addresses should follow a conventional, fixed pattern or whether it should be determined by the logic of the sequencing counter was considered to be of academic interest; the flight simulator was felt to be a sufficiently special-purpose device to justify either: 1. the programmer's care in assigning instructions to the addresses, or 2. the construction of a slow-speed translation device to convert binary forms into their machine equivalents. The capacity to specify a sequence of addresses out of the normal sequence as the result of a conditional or unconditional operation was incorporated in the sequencing counter, but this will not be discussed; it requires only the resetting of the sequencing counter to specify an arbitrary address.

Requirements

Further specification of the sequencing counter resulted from the following: 1. The sequencing counter must be capable of storing an address within its own logical circuitry. 2. The transition time from one address to the next must be minimized. 3. The equipment must be minimized. 4. The specification of a series of addresses need not follow a true binary sequence. The requirements are stated in the following definition of a particular type of sequencing counter which is acceptable for the purposes at hand.

Definition

A zero time counter is an m-stage register (where m is a specified integer) which advances through all its 2^m distinct states in response to a sequence of 2^m input counts in such a manner that, in the transition from any state k to its successor state $k+1$, all m storage cells in the register make the transition in one logical step and simultaneously to within the switching-time tolerance of the cells.

Methods Considered

Two methods for constructing a zero time counter were considered. The first technique involves the storage of additional information with each order in the instruction memory. Stored and read out with the kth instruction are sufficient bits to specify the address of the $(k+1)$th instruction. During the performance of the kth instruction these bits are held in a register in the control unit. At the conclusion of the kth instruction the register is interrogated and the $(k+1)$th instruction is specified.

The second technique involves an m-stage register with 2^m stable states. Each stage acts as a storage cell except a counting pulse occurs; in the latter case, the new state of each stage is determined as a logical function of the previous states of some or all of the other stages. The register advances through all its 2^m distinct states in response to 2^m counting pulses; each state is used to specify the address of an order in the instruction memory.

Logic Notation

The logic notation used in the remainder of the paper is illustrated in Fig. 1 by a typical, pulse-type, OR-AND-OR pyramid with pulse power amplifier and transformer.

N, greater than one, clock phases ($CP1$, $CP2$, ..., CPN) are assumed per pulse time, thus dividing a pulse time into N equal time intervals of duration $1/N$ units of time; $1/N$ is the standardized delay through a pyramid and its amplifier. The delay line in Fig. 1 is of length $(1-1/N)$ pulse times (nominal), a length which insures that an output pulse X will be recirculated and arrive at gate 2 one pulse time after it was generated; hereafter, such a delay pulse will be referred to by its corresponding small case letter x. Delay lines of nominal length $(1-1/N)$ and $(1-2/N)$ pulse times will hereafter be referred to by the symbols α and β respectively.

The regeneration gate 1 and the clock pulse inputs to AND gates 2 and 3 are used only for synchronizing and shaping purposes, and will be omitted from the logical circuit equations. In the equations multiplication represents AND and addition represents OR; a prime indicates negation, thus X' means "not X" or "X is not present." In the type of circuit discussed in this paper negation is realized by the application of a negative output pulse X^* (generated simultaneously with X) to an inhibit input of a gate denoted by a circle; thus X' also means "X^* is not present at an inhibit input."

The logical equation of Fig. 1 is therefore

$$(X, X^*) = xA + (B+C)D'$$

meaning that there will be a positive output pulse X (and simultaneously a negative output pulse X^*) if and only if both x and A are present, or either B or C or both (but not D^*) are present.

Previous Techniques

Both techniques mentioned in the section "Methods Considered" have been used previously in the construction of sequencing counters for digital computers. However, neither has resulted in a device which meets the present requirements. The purpose of this section is to outline previous methods of sequencing an instruction memory and discuss their limitations.

Paper 57-199, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted August 8, 1956; made available for printing September 13, 1957.

C. ELDERT and MORRIS RUBINOFF are with the Philco Corporation, Philadelphia, Pa.; H. M. GURK is with the Radio Corporation of America, Camden, N. J.; and H. J. GRAY, JR., is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.

This work was done as part of the project sponsored by the Special Devices Center, Port Washington, N. Y.

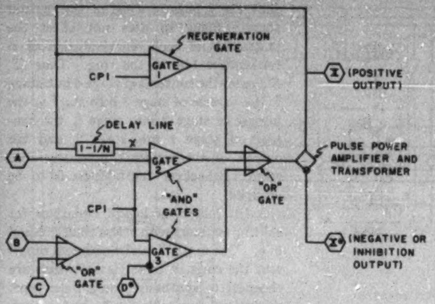


Fig. 1. Illustration of logic notation

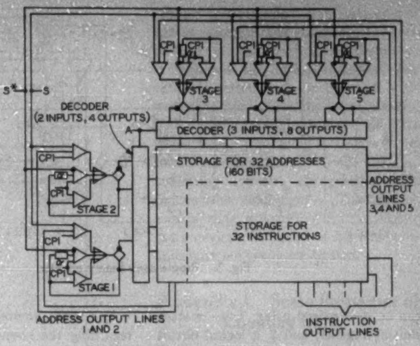

S*
S
CPI
STAGE 3
STAGE 4
STAGE 5
DECODER (2 INPUTS, 4 OUTPUTS)
A
DECODER (3 INPUTS, 8 OUTPUTS)
STORAGE FOR 32 ADDRESSES (160 BITS)
ADDRESS OUTPUT LINES 3,4 AND 5
STORAGE FOR 32 INSTRUCTIONS
STAGE 2
STAGE 1
ADDRESS OUTPUT LINES 1 AND 2
INSTRUCTION OUTPUT LINES

Fig. 2. Memory-sequenced conventional counter

MEMORY-SEQUENCED DEVICE

Fig. 2 illustrates a 32-word instruction memory, a 3-by-2 into 8-by-4 decoder, and a 5-stage dynamic register for sequencing the addresses. With each instruction are stored five additional bits (160 total) which are used to specify the address of the next order and which may be selected arbitrarily by the programmer.

Assume that the address of the instruction to be performed next in sequence is set up in the dynamic register. *A* gates the address into the decoder which specifies the proper memory location; both the instruction and its five additional bits are read out. The latter, which have been selected to specify the address of the next instruction, are gated by *S*, *S** into the dynamic register, replacing its previous contents.

Several comments are in order. First, the address of the instruction being performed is never available because the register is immediately advanced to the address of the next order. This may be remedied by: 1. reading the new address out of the memory into a second 5-stage dynamic register, and 2. gating the new address from the second register into the first at the conclusion of the current instruction. However, this alteration approximately doubles the amount of dynamic circuitry involved.

Second, the method requires the storage of $m2^m$ extra bits in the instruction memory in the general case. If the memory is a magnetic core matrix array, m additional matrices and their associated read out amplifiers and rewrite drivers are required.

Third, the technique is economically feasible only when the memory system and its associated selection equipment are already an integral part of the system. Furthermore, the speed of the sequencing device is limited by the access time of the memory. These last comments are not pertinent to the problem at hand but are made with reference to the general problem of constructing counters.

CONCERNING 2^m STABLE-STATE BINARY COUNTERS

Figs. 3 and 4 illustrate two previous alternatives to a memory-sequenced device. The examples taken are 5-stage counters with 32 stable states which follow a true binary sequence when triggered by *S*, *S**; both counters retain their states between applications of *S*, *S**. The decoder and instruction memory are omitted from the drawings since there are no feedback loops in the system.

The salient characteristic of the counter of Fig. 3 is that it operates on a ripple-carry basis. Application of (*S*, *S**) inverts the state of stage 1 and possible carries propagate through the remaining stages. The outputs of all the stages may not be read simultaneously until the *CP*1 following the application of (*S*, *S**), or four clock phases later. Furthermore, since the counter must drive pyramids in the decoder which are all clocked at the same phase (*CP*2), additional delay lines must be inserted in the outputs of stages 2, 3, and 4. (This last comment is also pertinent to the resetting of the counter to an arbitrary address when the drivers have a common clock phase.)

The 5-stage counter of Fig. 4 solves the ripple-carry problem by anticipating possible carries in combinational circuits. The transition from state k to state $k+1$ is made simultaneously in all stages at time *CP*1, and no delay lines need be inserted between the counter and the decoder. However, the logical complexity of the AND gates in the anticipatory carry pyramids is greater, the more significant the stage.

In summary, with the use of previous techniques it is possible to minimize equipment at the expense of ripple-carry time or vice versa; simultaneous minimization of both, however, does not appear to be feasible in a counter that follows a true binary sequence of states.

Shifting Counters

Returning to the definition of a zero time counter given in the section "Requirements," the anticipatory carry counter is obviously a special case, namely, a zero time counter which proceeds in true binary sequence. In an attempt to reduce equipment, other counting sequences were considered, with the requirement in mind of uniform minimum transition time. Since there are at least (2^m-1), different m-stage zero time counters, the following special class which offered hope of equipment minimization was investigated.

Consider such a closed sequence or ring of 2^m binary digits that any arrangement of m consecutive bits is represented once and only once in the complete ring. If, first, any m consecutive bits are sampled, and if the m consecutive bits displaced one digit to the left (right), etc., back to the starting point, are then sampled, all 2^m possible arrangements will have been encountered. Furthermore, every time the shift left (right) is made, only one new digit is encountered; the remaining $m-1$ digits were known from the previous number, and now simply appear shifted one place to the right (left) in the new number.

A counter operating on this principle

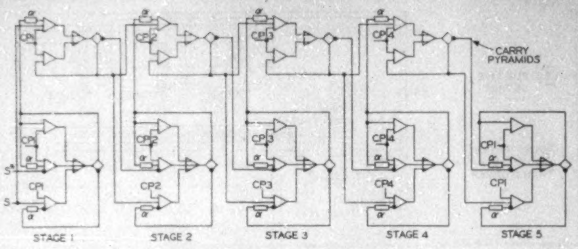


Fig. 3. Ripple-carry binary counter

can be mechanized using: 1. an m-stage shifting register, and 2. a logical circuit (or equivalent) which generates the next bit of the sequence as a function of the previous states of the m stages of the shifting register. To minimize equipment, it is necessary to design the simplest possible shifting register, and find that closed sequence which yields the least complex logical generator circuit. A method for generating the ring[1] and the number of different rings that exist[2] may be found in the literature. For m greater than 4, it is an overwhelming task to determine which ring possesses the simplest generator circuit. However, a few different rings have been developed for each value of m less than 7; the simplest rings of those investigated are presented in the section "Concerning 2^m Stable-State Shifting Counters." The section "Memory-Sequenced Device" illustrates a mechanization involving storage of the ring in the memory, a technique whereby generator complexity is of no consequence.

MEMORY-SEQUENCED DEVICE

Fig. 5 illustrates a 32-word instruction memory, a 3-by-2 into 8-by-4 decoder, and a 5-stage dynamic shifting register for sequencing the addresses. With each instruction is stored 1 additional bit (a total of 32) which forms part of the address of the next instruction.

Assume that the address of the instruction to be performed next in sequence is set up in the dynamic register. A gates the address into the decoder, which specifies the proper memory location; both the instruction and its additional bit are read out. (S, S^*) simultaneously gates the contents of stage 4 into stage 5, the contents of stage 3 into stage 4, . . . , and the bit from the memory into stage 1. The extra bit will have been selected once and for all by the programmer as the bit in the ring to the left of the m consecutive bits represented by the state of the shifting register before (S, S^*) occurred.

The same comments that were made regarding the memory-sequenced device of the section "Memory-Sequenced Device" are applicable here except that, first, only a single flip-flop need be inserted in the address output line in order to make available the address of the instruction which is being performed and, second, the method requires the storage of only 2^m bits.

CONCERNING 2^m STABLE-STATE SHIFTING COUNTERS

In those instances where it is not convenient to store the ring of 2^m bits, a logical generator circuit may be used as an alternative. Fig. 6 illustrates a 5-stage shifting register with a sequence generator pyramid to create the successive bits of the ring for insertion in stage 1 as a function of the previous states of the five register stages.

To simplify the hardware required to realize the logical equation for the generator circuit, it proves convenient to shift the primes of some of the register stages; doing so does not affect the uniqueness of each representation of m consecutive bits in the ring. Thus (S, S^*) gates the contents of stage 4 into stage 5, the inverse of stage 3 into stage 4, the inverse of stage 2 into stage 3, the contents of stage 1 into stage 2, and the output of the sequence generator into stage 1 whenever the address is to be changed.

A list follows of logical equations for shifting counters with m less than or equal to 6. The $m=2$, 3, and 4 counters represent the rings of those classes which are cheapest to synthesize using pulse-type, OR-AND-OR pyramids for logical gating. For $m=5$ and 6 many possible rings exist, and the counters listed represent the simplest of those that have been investigated (the $m=5$ counter is that shown in Fig. 6).

The logical equations for shifting counters with m greater than 6 were not investigated, although a ring for $m=12$ was obtained. It is a matter of speculation whether a shifting counter with m equal to 12, for example, may be more cheaply synthesized by providing the required generator circuit or by cascading counters of lower order. If an $m=12$ counter were formed from two $m=6$ counters, the equipment involved would be two generator circuits plus an additional pyramid to sense a particular state (say, all zeros) of the low-order counter. Whether this amounts to more or less equipment than the $m=12$ generator circuit is not known.

The counting sequence of the $m=5$ shifting counter of Fig. 6 is as given on the following page:

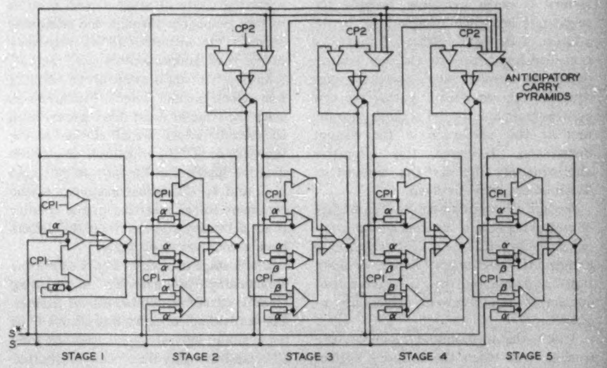


Fig. 4. Gated binary counter

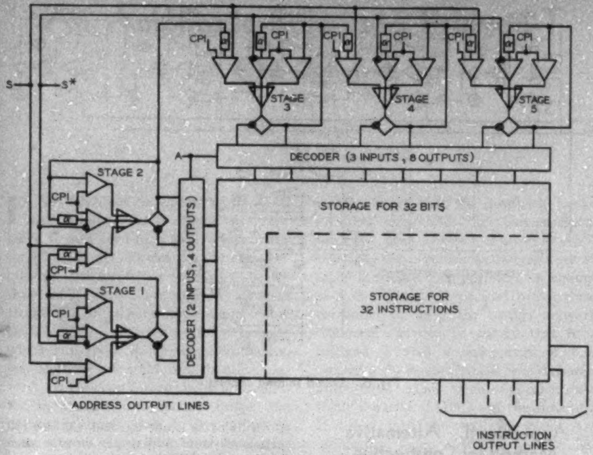


Fig. 5. Memory-sequenced shifting counter

Stages				
1	2	3	4	5
0	0	0	0	0
0	0	1	1	0
0	0	1	0	1
0	0	1	0	0
1	0	1	0	0
1	1	1	0	0
1	1	0	0	0
1	1	0	1	0
1	1	0	1	1
0	1	0	1	1
1	0	0	1	1
1	1	1	1	1
0	1	0	0	1
1	0	0	1	0
0	1	1	1	1
1	0	0	0	1
1	1	1	1	0
1	1	0	0	1
0	1	0	1	0
0	0	0	1	1
1	0	1	1	1
0	1	1	0	1
1	0	0	0	0
0	1	1	1	0
0	0	0	0	1
1	0	1	1	0
1	1	1	0	1
0	1	0	0	0
0	0	0	1	0
0	0	1	1	1
1	0	1	0	1
0	1	1	0	0
Repeat				

Logical Equations

$m=2$ (4 states)
$A=Sg+S'a$
$B=Sa'+S'b$
$g=b$

$m=3$ (8 states)
$A=Sg+S'a$
$B=Sa'+S'b$
$C=Sb+S'c$
$g=[C(A+B)+A'B'C']$ delayed by β

$m=4$ (16 states)
$A=Sg+S'a$
$B=Sa'+S'b$
$C=Sb+S'c$
$D=Sc+S'd$
$g=[D(A+B+C)+A'B'C'D']$ delayed by β

$m=5$ (32 states)
$A=Sg+S'a$
$B=Sa+S'b$
$C=Sb'+S'e$
$D=Sc'+S'd$
$E=Sd+S'e$
$g=[AE'(B+C)+A'E(B+D)+B'C'E+B'CD'E'$ delayed by β

$m=6$ (63 states)
$A=Sfg+Sf'g'+S'a$
$B=Sa+S'b$
$C=Sb+S'c$
$D=Sc+S'd$
$E=Sd'+S'e$
$F=Se'+S'f$
$g=[A'(B+C+D)+AB'D'E+BC'DE+B'C(D+E)]$ delayed by β

Note that for $m=6$ the A stage is logically gated and this counter is therefore not a simple shifting counter. The choice of this mechanization simplified the generator circuit g.

Conclusions

By relaxing the restriction that the sequence of addresses follow a true binary sequence, methods have been presented whereby the instructions in a 1-address digital computer may be sequenced at the pulse repetition rate of the machine. The methods offer the advantage of equipment minimization whether a memory-sequenced device or a 2^m stable-state counter is employed. Synthesis of the former mechanism is straightforward regardless of the size of the memory; synthesis of the latter mechanism has not been attempted for counters with more than 64 stable states. In the event that a memory contains more than 64 instructions, cascading of low-order counters is suggested.

Shifting counters are possible with radices other than binary. The implications of using, in a shifting counter, higher radix devices composed of bistable subunits have not been investigated.

Appendix I. Some Properties of Counting Rings

Consider a sequence $a_0, a_1, \ldots, a_n$, of the following type: Given $m\geq 1$

1. $a_i=0$ or 1 for $i=0, 1, \ldots$
2. $a_i=a_{i+2^m}$ for $i=0, 1, \ldots$
3. $a_i, a_{i+1}, \ldots, a_{i+m-1}=a_j, a_{j+1}, \ldots, a_{j+m-1}\rightarrow i\equiv j$ (modulus 2^m)

Only consecutive subsequences of this sequence will be considered: thus by "subsequence" is meant consecutive subsequence.

By items 1, 2, and 3 it is seen that the sequence a_i may be thought of as a cyclic sequence of 2^m 0's and 1's (i.e., the 2^m digits are arranged in a ring) so that no two subsequences of length m are exactly the same. Such a sequence is called a "full counting sequence" or "full counting ring." By a "truncated counting sequence" is meant such a cyclic sequence of $k<2^m$ digits that no two sequences of length m are exactly the same.

Mainly the full counting sequences are of interest. All such sequences have the property that there must be exactly

$$\frac{2^m}{2^k}=2^{m-k} \text{ subsequences of 1's}$$

of length k and 2^{m-k} subsequences of 0's of length k for $1\leq k\leq m$, and no such sequences for $k>m$.

By a maximal subsequence of 1's (or 0's) of length k is meant a subsequence of 1's (0's) of length k which is not a subsequence of another subsequence of 1's (0's) of length $k+1$. With this definition and the foregoing property the following may be proved.

Theorem 1: For a given m, there is a unique function $M(k)$, $k\geq 1$, in which $M(k)$ is the number of maximal subsequences of 1's (0's) of length k in any full counting sequence.

To prove theorem 1 the following lemma are needed:

Lemma 1

$$m+\sum_{j=1}^{m-2} 2^{j-1}(m-j-1)=2^{m-1}$$

for $m\geq 1$ when $\sum_{j=1}^{m-2}=0$

if $m-2<0$

Proof: The lemma holds for $m=1, 2$. Assume it holds for $m=A$:

$$A+\sum_{j=1}^{A-2}2^{j-1}(A-j-1)=2^{A-1}$$

Multiplying by 2 yields

$$2A+\sum_{j=1}^{A-2}2^{j}[A-(j+1)]=2^{A}$$

$$A+1+A-1+\sum_{j=2}^{A-1}2^{j-1}(a-j)=2^{A}$$

$$(A+1)+\sum_{j=1}^{(A+1)-2}2^{j-1}[(A+1)-j-1]=2^{(A+1)-1}$$

By induction, the following assertion results: There is one maximal subsequence of 1's of length m. There are two subsequences of this subsequence of length $m-1$. Thus no more subsequences are needed of 1's of length $m-1$. There are only three subsequences of the m-long subsequence of 1's of length $m-2$.

Continuing in this manner by induction: Assume for $0<k\leqslant s\leqslant m-2$ there are 2^{k-1} maximal subsequences of 1's of length $m-k-1$, no such sequences of length $m-1$, and one of length m.

For a sequence of 1's of length r, there are $r-t+1$ subsequences of 1's of length t, $t\leqq r$. Thus there must be the following number of maximal subsequences of 1's of length $m-s-2$:

$$\frac{2^m}{2^{m-s-2}}-\left\{M+1-(m-s-2)+\sum_{k=1}^{s}2^{k-1}\right.$$

$$\left.[m-k-1+1-(m-s-2)]\right\}$$

$$=2^{s+2}-\left[s+3+\sum_{k=1}^{s}2^{k-1}(s+2-k)\right]$$

$$=2^{s+2}-\left[s+3+\sum_{k=1}^{s}2^{k-1}(s+3-k-1)\right]$$

$$=2^{s}(s+3-s-1-1),\text{ by lemma 1}$$

$$=2^{s}$$

Thus, there are 2^s maximal subsequences of 1's of length $m-(s+1)-1$.

By induction, $M(k)$ uniquely becomes

$$M(k)=\begin{cases}0 & k>m\\ 1 & k=m\\ 0 & k=m-1\\ 2^{m-k-2} & m-2\geqslant k\geqslant 1\end{cases}$$

Similarly, $M(k)$ proves also to be the unique number of maximal subsequences of 0's in any full counting sequence.

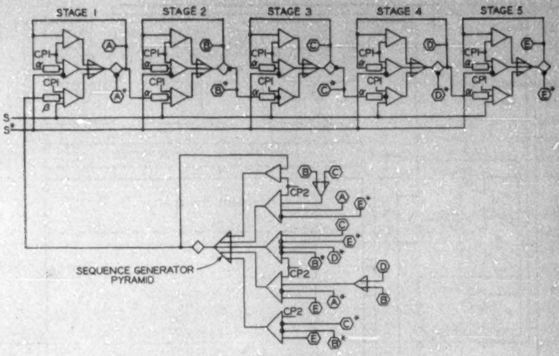


Fig. 6. Gated shifting counter

Appendix II. Alternative Method for Constructing Counting Rings

The following is a method for constructing a cyclic sequence which under certain circumstances is a full counting sequence and otherwise is a truncated counting sequence. (The method has been successful for all values of m tried to date, including $m=12$.) The process is described inductively.

The various m-long subsequences may be thought of as the numbers from 0 to 2^m-1 represented in binary notation.

1. Let the first m digits in the sequence be 1's; in binary notation, the number 2^m-1.

2. Let 0 be the $(m+1)$st digit in the sequence.

3. Suppose one has proceeded through the kth step. If the last digit in the sequence after the kth step (called the "kth step sequence") is an i ($i=0$ or 1), then the $(k+1)$st step consists of adding an i to the sequence, provided:

   a. The number of consecutive i's at the end of the newly formed $(k+1)$th step sequence is r; then $M(r)$ maximal i-sequences of length r have not previously appeared in the sequence, and

   b. The number represented by the last m digits of the $(k+1)$th step sequence has not been represented in such form by any previous subsequence of m digits.

4. If 3*a* or 3*b* is not satisfied, the $(k+1)$st step consists of adding the inverse digit $1-i$ to the kth step sequence provided 3b now holds for the new $(k+1)$th step sequence.

5. If neither of the steps 3 or 4 is possible for the $(k+1)$st step, no new digits are added to the sequence, but rather:

   a. The beginning and end of the kth step sequence of $i=0$ are joined and thus a cyclic sequence is obtained, or

   b. If $i=1$ and r is the number of 1's ending the kth step sequence, the sequence cyclic is made by considering these r 1's as the first r of the initial sequence of m 1's.

The following theorem can be proved but the proof is too long for inclusion here.

Theorem 2: If in using the foregoing procedure the sequence cyclic must be made by method 5a, a full counting sequence is obtained; if method 5b must be used, a truncated counting sequence results. Thus whenever it is no longer feasible to proceed beyond a 0 with the method described in either 3 or 4, a full counting sequence will be obtained; when not beyond a 1, a truncated sequence will result.

References

1. A Problem in Arrangements, M. H. Martin. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 40, 1934, pp. 859–64.

2. Circuits and Trees in Oriented Linear Graphs, N. G. deBruijn, T. van Aardenne-Ehrenfest. *Simon Stevin*, Groningen, Netherlands, vol. 28, 1950–51, pp. 203–17.

An Amplitude-Stabilized Bridged-T Oscillator

KURT ENSLEIN
MEMBER AIEE

IN AN ANALOG computer application which used several nonlinear elements, there arose a need for a sinusoidal oscillator of 2,500 cps (cycles per second) frequency, with frequency stability of better than 0.5%, amplitude stability greater than 0.2%, and a harmonic content of less than 0.2%. There are very few circuits which can reconcile these requirements, one of these being the Meacham oscillator.[1] If the requirement is for an oscillator in the audio-frequency range, the Meacham circuit loses its attractiveness, since quartz crystals in this frequency range are rather cumbersome. If an inductor and a capacitor are substituted for the crystal, it is found that some other circuits have better inherent phase magnification.

The problem of good frequency stability is usually solved concurrently with the problem of good waveform. One of the best circuits from this standpoint is the bridged-T oscillator,[1,2] with the use of the resonant network of Fig. 1. However, the conventional circuit has the inherent weakness that its amplitude stability is a function of the Q of the resonant series circuit. In the audio-frequency range this Q is, for all practical purposes, determined by the inductor. The Q of inductors is not a constant, but varies with temperature. (The effect of a change in Q is to unbalance the network. This changes the feedback factor and hence the oscillation amplitude; see the section "Control Network Characteristics.")

As an illustration of this effect, consider an oscillator with an open loop gain $A=600$. For a reasonable amplitude of oscillation, the closed loop gain $a=2$, which means that the feedback factor $\beta=2/600$ or 0.0033. Let this oscillator use a coil, the Q of which changes by 0.5% when the temperature is changed from 30 to 60 C (degrees centigrade). A change in R of 2% will be required with typical constants to keep the bridge network balanced. It can be shown that in a bridged T with a component of this nature, a change in R of 2% would change β to 0.011. This would approximately triple the output voltage of the oscillator. If β decreased, the oscillator would stop oscillating, since a would become less than 1, or even negative. Then, the need for amplitude stabilizing a bridged-T oscillator obviously is indicated.

Method of Approach

A number of methods[1-3] have been developed in the past for the amplitude stabilization of class A oscillators. These may be divided into two categories: thermistor control and rectified-signal feedback.

The thermistor method, while quite practical, cannot produce stabilization ratios, i.e., $[f(T)$ with compensation$]/[f(T)$ with compensation$]$ greater than about 20, due to inherent limitations in the temperature coefficients of thermistors. The second method may be applied in two ways: either one may control the amplitude of oscillation directly by changing the gain in the oscillator loop, or one may pass the variable oscillator signal through a variable gain amplifier, the latter being controlled by its own rectified output signal. In the first case, only relatively small stabilization ratios can be obtained in single circuits and, in the second case, variable μ tubes inherently also have variable μ for the signal, and therefore introduce harmonic distortion. The same problem exists in the control of an oscillator by changing the μ of all of the tubes in the loop. Still another method uses the frequency-determining network itself to control the positive feedback signal, without at the same time influencing the frequency of resonance, as shown in schematic form in Fig. 2. In this method, the network, in addition to its normal input and output terminals, also has a "control" terminal. When the proper potential is applied to this terminal, the "balance" conditions of the network may be made to change in response to the control potential. Some of the networks that might be used also have the property of phase inversion for the fundamental (oscillation) frequency, and no phase inversion for harmonics. The loop $A_1(\omega)$ can therefore be a positive feedback loop, and can oscillate if its gain ≥ 1.

The output signal is also rectified, and compared to a reference level. The difference between reference and rectified signal is amplified in A_2 and applied to the control terminal of the network in such a sense as to minimize changes in the magnitude of the output voltage. If the control gain of the network and A_2 can be made large enough, it should be possible to obtain any desired degree of amplitude stabilization of such an oscillator.

Control Network Characteristics

References 2, 4, and 5 show that one of the better compromises for a selective network to be used in an oscillator is the bridged T. This network, shown in Fig. 2, has several interesting properties.

First, its frequency response is expressed by

$$A(w)=\frac{1}{1-j\dfrac{\rho}{\rho^2-1}\dfrac{2}{Q_0}} \tag{1}$$

where

$$\rho=\frac{w}{w_0}=\frac{\text{operating frequency}}{\text{null frequency}}$$

For Q of the order of 10, the network has a very small loss at the second and higher harmonics of the fundamental frequency. This means that if the network is used in a negative feedback loop, β approaches unity at the second and higher harmonics, and distortion components are consequently reduced to very small values.

Second, the network can produce zero output under the following conditions:

$$Q_0=w_0 2RC \tag{2}$$

where

$$Q_0=\frac{w_0L}{2}$$

and

$$w_0^2=\frac{2}{LC} \tag{3}$$

An analysis to determine the response of the network at w_0 as R is changed

---

Paper 57-76, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 16, 1956; made available for printing August 15, 1957.

Kurt Enslein, formerly with the University of Rochester, Rochester, N. Y., is now with Brooks Research, Inc., Rochester, N. Y.

The original suggestion of W. F. Coombs, Jr., to use the bridged-T oscillator for the low distortion requirement should be acknowledged, as well as the untiring efforts of Anthony Formicola in building, improving, and "trouble-shooting" the circuit and making all measurements reported in this paper. The work was supported by the Atomic Energy Commission.

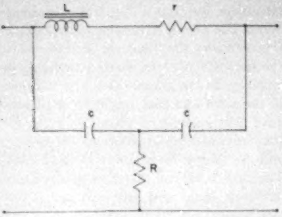


Fig. 1. Bridged-T network

from the value required for balance shows that for $R > R_0$, where R_0 = resistance required to balance the network, the network produces an output voltage proportional to the change in R, with the same phase as the input voltage. If $R < R_0$, the output is still proportional to the change in R, but the output voltage is now 180 degrees out of phase with the input voltage. All this is accomplished without influencing the response of the network at the harmonics.

In the oscillator of Fig. 2, it is then possible, by making R slightly larger than R_0, to use a bridged-T network for both positive and negative feedback, without prejudicing the frequency stability of the network in any way, or increasing the attenuation of the network at harmonics of the oscillator frequency.

Third, a change in Q, which is equivalent to a change in r, may be compensated by a corresponding change in R, without affecting the resonance frequency. There is, however, a slight change in phase shift with frequency, due to the basic change in Q. This will result in a very slight change in oscillation frequency. The salient fact, however, is that, by adjusting R, the attenuation of the network may be kept constant as Q is changed.

This latter characteristic is of fundamental importance when temperature changes are considered. The temperature changes are almost always accompanied by variations in Q, and therefore can cause changes in the amplitude of oscillation of the system.

It is then evident that, if a variable resistor is substituted for R, a means for keeping the oscillator amplitude constant has been obtained. Such a resistor[8] is the dynamic plate resistance of a vacuum tube. If the proper tube is chosen, and operated in a region where it exhibits a large slope of plate resistance versus plate current, it is conceivable that the network might have some control gain greater than one.

The Network

CHOICE OF INDUCTOR

The oscillator in question is intended to operate at 2.5 kc, over a temperature range of 30 to 60 C. From the standpoint of frequency stability and waveform purity, a high Q is desired, preferably greater than 10. At the frequency in question, in order to reduce the coil to a reasonable size, a fairly high-permeability (μ) core is desirable. However, in general, the higher the μ of the core (due to its lower Curie temperature), the higher the change in μ with temperature, and therefore the greater the change in response of the network with temperature. Also, when the Q is high enough, a major part of the losses reside in the resistance of the wire. The resistance also changes with temperature and may be compensated to a fair degree by means of thermistors.

Table I shows temperature coefficients for two typical inductors that are usable in the oscillator in question. Both have a nominal inductance of 2.5 henrys at 1 kc. It may be seen that the temperature coefficient of inductance of the two units is not materially different, but the temperature coefficient of Q (as indicated by the required change in R to rebalance the network) is much greater for the type *MQA-14*. This latter inductor has a Q of the order of 90 at 2.5 kc, whereas *MQL-1* has a Q of only 6.35. Since a reasonably low distortion level may be obtained with a Q of 6.35, and since amplitude stability is such an important consideration in this case, the *MQL-1* unit was chosen. The particular sample used had an inductance of 2.92 henrys at 2.5 kc. Practically all the losses were due to the core, which probably accounts for the smaller temperature coefficient of Q.

With the hindsight possible after the completion of this project, it is not very clear whether the choice of the low-Q coil was a wise decision. While the high-Q coil has a higher temperature coefficient of Q, it seems plausible that it would have been possible to increase the loop gain to make up for this larger variation. This is especially so in view of the methods of obtaining R, described in a later section. The use of a higher-Q coil would of course have entailed a lower harmonic distortion.

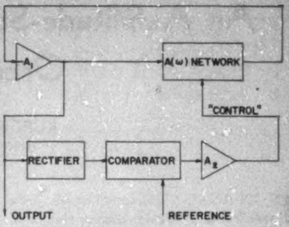


Fig. 2. Block diagram of system

GAIN

Fig. 3 shows a computed gain characteristic for a bridged-T network with components somewhat different from the ones used in this application. It will be noted that the phase reversal mentioned earlier occurs. In the network used for this application, the measured control gain is approximately 60.

CHOICE OF CONTROL TUBE CIRCUIT

Several conflicting requirements are imposed on the control tube. On the one hand, the circuit requires a dynamic plate resistance of approximately 70 kilohms; on the other hand, its d-c resistance should be rather high, so as to permit the use of a large series-dropping resistor. This resistor will be in parallel with the dynamic resistance of the tube, and should therefore be large, in order not to reduce the gain of the system appreciably. Also, one would like to obtain as large a change in plate resistance with grid voltage as possible; yet there should be no appreciable nonlinearity of plate resistance versus plate voltage over the desired operating region.

This latter point probably requires explanation. When a resistor is used as R in the bridged-T network, no consideration need be given to its voltage character-

Table I. Temperature Characteristics of Two Inductors

Inductor	T, Degrees Fahrenheit	$\Delta I\%$	$\Delta I\%$, Degrees Fahrenheit	$\Delta R\%$	$\Delta R\%$, Degrees Fahrenheit
MQL-1	79				
	125	−0.56	−0.12	+1.66	+0.036
	150	−0.72	−0.010	+3.04	+0.043
MQA-14	79				
	125	−0.500	−0.011	−10.05	−0.218
	150	−1.35	−0.019	−17.5	−0.246

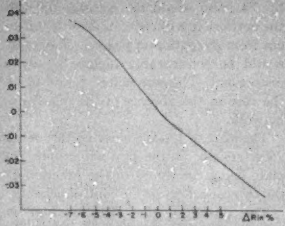


Fig. 3. Computed input-output characteristic of bridged-T network

istic, since the resistor is usually far more linear than the other components in the bridged T. However, when the dynamic plate resistance of a triode is substituted, the possibility exists that plate resistance versus plate voltage is a nonlinear function. Since half the oscillator signal appears across R, the control tube plate resistance should vary linearly when its plate voltage is changed by half the oscillator output voltage. In this particular case, the desired output of the oscillator is 10 volts, peak to peak. The incremental voltage range of interest for plate resistance linearity is then 5 volts.

The question now arises as to what type of tube should be used. Since the requirement is for dynamic plate resistance of approximately 70 kilohms, a triode is clearly indicated. Then the question becomes whether one should use a low μ or a high μ triode. Reference 7 shows that the μ of a triode is quite independent of plate potential, but that gm varies as

$$(E_g - E_{p/\mu})^{1/2} \tag{4}$$

Since $R_p = \mu/gm$, plate resistance will vary with plate voltage unless some precautions are taken. The most obvious solution is to use a small plate voltage and a high-μ tube, so that the second term in equation 4 is minimized. This process is limited by the minimum plate voltage that may be used on a given tube before excessive grid current is obtained so that, by and large, the second term of the equation is usually approximately equal to the first term. However, the effectiveness of the grid in screening the anode-grid space from the cathode-grid space is usually better in high-μ triodes. Therefore one may expect fewer nonlinearities to be encountered with a high-μ triode than with a lower μ tube.

A second method would consist of keeping the second term in equation 4 constant, by choosing a tube for which μ changes linearly with E_p. This is basically a very difficult method, and does not represent a practical approach unless a special tube were to be constructed for the purpose.

A third method consists of introducing a cathode resistor. The effective plate resistance of a triode with a resistor in the cathode to ground lead is

$$R_p \text{ eff} = R_p + (\mu + 1)R_k \tag{5}$$

The second term in equation 5 may be made dominant, which in turn makes the second term in equation 4 less important. The only problem for this particular design is that of obtaining 70 kilohms. For the second term of equation 5 to be dominant, it must represent most of this 70 kilohms, and therefore R_p must be rather small. This implies either rather high-current operation of the triode, or a basically low R_p tube, which usually also means low μ.

Unfortunately, for the value of $R_{p\ eff}$ desired, an improvement in linearity of only approximately 2.5 times may be obtained with presently available tubes. Due to the cathode resistor, the gain of the control tube is necessarily less, since the cathode resistor causes a form of degeneration. On the other hand, the scheme of Fig. 4 may be used to dispose of the problem. In this method of operation, $R_{p\ eff}$ is varied by changing R_k, which change is accomplished by changing the transconductance of V_2. While it might be surmised that this circuit is compensated for filament potential changes, much as a differential d-c amplifier, this is not entirely the case, since the compensation is effected by virtue of a change in R_p. With typical tubes, the control gain $[(dR_{p\ eff})]/(dE_{g2})$ is of the order of 20 kilohms per volt, whereas a simple triode such as a *12AX7*, can produce changes of 70 kilohms per volt.

Tests of various tubes (*12AX7*, *12AU7*, *5965*, *6J6*, *6463*) reveal that the best compromise of linearity and slope for this design may be obtained with the *12AX7* operating at approximately 0.5 milliampere, and at least a 100-volt plate-to-cathode potential.

There remains one consideration in the choice of the tube type, if the oscillator is to be used at moderately high frequencies. Due to the Miller effect in the control triode, it is possible to obtain unstable operation if enough oscillator signal is allowed to appear at the grid of the triode. The remedy for this difficulty is to make the impedance in the grid low enough at the oscillator frequency. This has been accomplished by means of a by-pass capacitor from grid to ground.

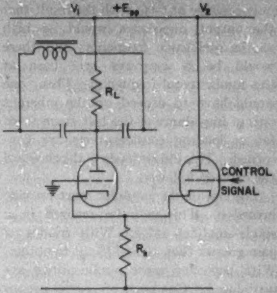


Fig. 4. Variable R circuit

The A-C Loop

The oscillator loop for this design should have several characteristics which are easily defined. It should have high gain, in order that the harmonics will be largely fed back. Yet inherent harmonic distortion should be low, so that the negative feedback of the harmonics need not work against too severe odds. The output tube should operate at such a level that it can deliver 10 volts peak-to-peak into an impedance of 10,000 ohms. This means a fairly high-level output stage.

Finally, the loop phase shift, excluding the frequency-determining network, should be 180 degrees and should vary very little with frequency near the oscillation frequency. In this case, this means that coupling and by-pass time constants should have a period long compared to 400 μseconds. Actually, in order to reduce the effects of stray 60 and 120 cps (introduced by filaments, etc.), it is desirable that the open loop gain be high at these low frequencies, so that essentially unity negative feedback will be available for these components.

Referring to Fig. 5, the a-c loop consists of the network described in the previous section, V_1 and V_2. This particular circuit was arrived at from the following considerations: 1. The *6Y6-G* was chosen as a cathode follower to give low output impedance and because of its relatively low μ. This latter feature gives a long grid base, which in turn means that the signal swings over a minute fraction of it. While the cathode-follower action effectively increases the base, lower power or higher μ tubes still produced distortion too high to be tolerated here. While an anode follower[8] would do as well (except for at least three times the output impedance), a simple amplifier would not, as there would be negative feedback at $f \neq f_0$, but only positive at $f = f_0$, with the result that the output impedance would be high at the oscillation frequency, and there would be a narrower grid base at the fundamental frequency. Then, one would have to depend on the inherent output impedance of the last stage in so far as loading considerations are concerned. This clearly leads to the choice of the cathode follower.

The rest of the amplifier must provide inversion. This may be realized in a single-amplifier stage. With triodes, a gain greater than 70 is difficult to obtain. With pentodes, more nonlinearities are introduced. One may, of course, use an odd number of stages greater than 1, but the complications, cumulative distortion, etc., become rather cumbersome.

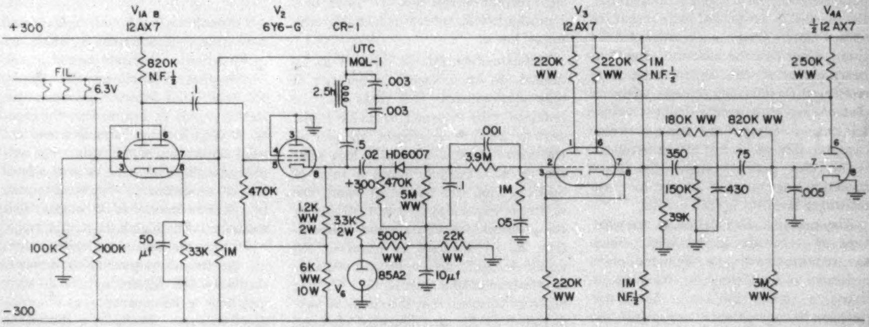


Fig. 5. Over-all circuit

As a compromise, the cascode circuit is then chosen. This gives a gain of approximately 700, without substantial nonlinearities. The cascode is very much in need of a high impedance load, due to its very high output impedance. This is another good reason for the choice of the cathode follower output stage.

The Control Loop

The control loop, Fig. 5, includes, besides the control network and the a-c loop, V_3 and V_4 and *CR-1; CR-1* is a Hughes Aircraft Company silicon junction diode, which serves the purpose of peak-rectifying the oscillator output. This diode was chosen in preference to a vacuum diode because of its lack of filament. Since its back resistance compares favorably with a vacuum diode, no loss in rectification efficiency has been incurred, except perhaps for the high knee of the silicon diode. The lack of filament eliminates any problems of 60- and 120-cps modulation at this stage, and also circumvents the variability of filament potential, which would have the effect of introducing an error at the input of the system, and therefore affect the output amplitude.

The output from the diode is bucked against a fixed d-c voltage derived from a gaseous discharge tube of good temperature stability. (A Zener diode is another good possibility.) The resultant voltage is applied, through a stabilizing network, to the grid of $V3_a$, which is one tube of a differential amplifier, with a gain of the order of 30. The other side, $V3_b$, has impressed on its grid a voltage which is proportional to the difference between the positive and negative power supplies. This connection has the effect of compensating for changes in the supply potentials, in order to permit the use of less well regulated power supplies. Later data will demonstrate the effectiveness of this arrangement. The differential amplifier is used in preference to a single triode in order to obtain compensation for filament potential changes. These filament potential changes, if not compensated, would have the effect of introducing a spurious input voltage to the amplifier, and therefore would appear as an error signal, even when no error exists.

The output of $V3_b$ is coupled to a parallel-T network, tuned to the oscillation frequency. The output of the network is coupled through a d-c divider to the control tube $V4_a$.

The null filter is needed to reduce the oscillation frequency component on the grid of the control tube, for oscillation voltage at this point would be amplified by $V4_a$, and would be reintroduced into the network in a phase considerably different from the optimum phase. The result would be considerably higher harmonic distortion. This was indeed found to be the case in the absence of the parallel-T network.

The plate of $V4_a$ is connected through a *250K* resistor to the positive supply, its plate being connected to the junction of the two capacitors in the control network.

The choice of the *250K* resistor is influenced by two considerations: its value should be large compared to the plate resistance of the control tube so that it does not shunt it appreciably, and its value should also be such that, with the current required through the control tube to establish the proper resistance, it does not reduce the plate voltage to a value so low that appreciable grid current will flow.

Since about 0.5 milliampere flows through the control tube, the plate voltage here proves to be 175 volts, and the ratio of the plate resistor to the plate resistance is about 3.5 to 1. This seems a reasonable compromise.

The point that the plate resistor should not appreciably shunt the control tube is debatable, for if a smaller value is used the control tube must work in a higher plate resistance region, and it is entirely possible that in this region, the over-all control gain may actually be higher. There is also the possible advantage that since most of the resistance would be ohmic, nonlinearities in the plate resistance characteristic would be less likely to affect the linearity of the oscillator. This aspect of the problem has not been thoroughly investigated.

Performance

Frequency Stability Versus Temperature

The frequency stability was measured with temperature as a variable. The range of measurement was from 24 to 66 C in all these tests. From the data in Table I, one would expect a maximum frequency change of 0.38%. However, in the actual circuit, the peak-to-peak deviation of frequency was 0.62%, the average deviation less than 0.05%. Most of the large deviations occurred during the warm-up period, and are probably due to the heating of the wire in the coil. The core, having a much longer time constant, did not heat as rapidly. That the average change is less than expected must be attributed to a fortunate choice of other frequency-determining components (the two capacitors in the bridge). Also, the amplitude stabilization of this circuit operates by virtue of a slight phase shift, which in this case is in a direction such as to aid the frequency stability.

Frequency Stability Versus Supply Voltage Changes

Changing the supply voltage by ±3.3% produced an undetectable change in frequency. The level of detection here is one part in 2,500.

Distortion

At a 10-volt peak-to-peak output, the second harmonic content was 0.08% and the third was 0.02%; further harmonics were negligible compared to the first three. Total harmonic distortion was therefore of the order of 0.1%.

Amplitude Stability Versus Temperature

Using the same range of temperature variation as for the frequency stability measurement, the peak-to-peak amplitude variation was 0.5%. After the coil had stabilized at the new temperature, the deviation could not be measured, which means that it was less than 0.02%.

Amplitude Stability Versus Supply Voltage Variations

Before the network in the grid of $V3$, was inserted, a 3.3% change in one of the supply potentials produced a 5% to 6% change in output level. After the network was inserted, a −3.3% change in the positive supply changed the output by −0.25% and a −3.3% change in the negative supply caused a −0.16% change in the output. If both supplies are then varied simultaneously, which is usual for line voltage fluctuations, the net change for a 3.3% change in supply voltages will be −0.09%.

The filament voltage on all tubes was changed by ±10%. The observed effect on the output is a change of less than 0.1%.

Other Possible Applications

Due to the amplitude-regulating characteristics of this oscillator, a number of other uses suggest themselves. For example, it would seem very logical to use the oscillator as an amplitude-modulated device, of very low distortion. For example, this can easily be done by superimposing an a-c signal on the d-c signal on the grid of the control tube.

It is also possible to substitute a saturable reactor for the coil used here, in order to provide a frequency-modulated oscillator. In this case, since such reactors usually change in Q along with inductance, due to the variable saturation of the core, and since a change in frequency will mean a required change in R, the circuit would tend to keep the amplitude stable independently of frequency, and without affecting the harmonic distortion content. In short, this approach is applicable to cases where, for one reason or another, the output amplitude of an oscillator changes in response to various disturbances, and where it is desired to keep the amplitude of the oscillator constant.

References

1. **Vacuum-Tube Oscillators (book), W. A. Edson. John Wiley & Sons, Inc., New York, N. Y., vol. 9, 1953, pp. 126–57.**
2. **Waveforms (book), B. Chance, et al. McGraw-Hill Book Company, Inc., New York, N. Y., MIT Radiation Laboratory Series, vol. 19, pp. 115–23.**
3. **Theory of Amplitude-Stabilized Oscillators, P. E. Aigrain, E. M. Williams.** ***Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, no. 1, Jan. 1948, pp. 16–19.**
4. **Bridged-T and Parallel-T Nul Circuits for Measurements at Radio Frequencies, W. N. Tuttle.** ***Ibid.*, vol. 28, no. 1, Jan. 1940 pp. 23–29.**
5. **Vacuum-Tube Amplifiers (book), G. E. Valley, H. Wallman. McGraw-Hill Book Company, Inc., MIT Radiation Laboratory Series, vol. 18, 1948, pp. 384–91.**
6. **Wide-Range Deviable Oscillator, M. E. Ames.** ***Electronics*, New York, N. Y., vol. 22, May 1949, pp. 96–100.**
7. **Radio Engineers' Handbook (book), F. E. Terman. McGraw-Hill Book Company, Inc., 1943, pp. 305–08.**
8. **Anode-Follower Derivatives, A. W. Keen.** ***Wireless Engineer*, London, England, vol. 30, no. 1, Jan. 1953, p. 5.**

New Printing Demand Meter

P. V. TERRY
NONMEMBER AIEE

Synopsis: A new printing demand meter has been designed. Several innovations have resulted in a device which has higher accuracy, lower cost, and smaller size than previous printing demand meters. Some of the novel features of the new improved design are discussed.

THREE TIMES the amount of electric energy was sold in 1957 as in 1947. Electric energy usage will almost triple again by 1970. These figures[1] are indicative of the rapid development of the electrical industry. It is growing at a far greater rate than the total population or the number of electric energy users.

Likewise, in the metering of kilowatt demand the energy used during the demand intervals has been increasing. On large connected loads where an impulse-operated printing demand meter is used to obtain a permanent demand record, each impulse can represent an extremely large number of kilowatts which in some cases may be considered excessive for good metering practice. A printing demand meter capable of accepting an increased number of impulses per demand interval became necessary to reduce the kilowatt value of each impulse. Impulses which up to now had been lost during the print and reset operation of the demand meter are important and should be recorded to improve the accuracy of demand measurement and customer billing.

The new printing demand meter described in this paper was designed to provide the needed long-life meter which could more accurately measure and permanently record electrical demand information; see Fig. 1. Accuracy compatible to that of the watt-hour meter, which acts as the primary sensing device and impulse generator for the printing demand meter, is a design feature of the new printing demand meter.

The purpose of this paper will be to describe some of the unique design approaches which were employed to result in a more accurate, compact, and lower cost printing demand meter. The portions of the over-all design to be discussed are: 1. The printing mechanism. 2. The record paper. 3. The impulse accumulator and governor. 4. The reset mechanism for the running demand indicator.

The Printing Mechanism

In the past, printing has generally been accomplished by utilizing a solenoid to accelerate a printing arm and platen. The kinetic energy, which the solenoid transmitted to the printing arm, was used to obtain a printed record on a paper tape by impact of the platen with carbon tape over record tape over the raised numerals on the counting wheels. A variation in the clarity and density of the record occurs when this system is used since the pull-in speed of the solenoid is a function of voltage which may vary up to 10% in some locations. The pull-in speed also depends upon the portion of the cycle in which energization occurs. To overcome this difficulty, a spring-driven platen arm actuated by a motor-driven cam is used to provide constant printing arm energy resulting in uniform printing density and clarity; see Fig. 2.

In the past, the necessary adjustments to line up the face of the platen with the numeral faces have been provided. By using a frictionally restrained ball joint to secure the platen to the printing arm, the assembly becomes self-adjusting. A further advantage is gained with this construction because of the ease with which an old platen may be snapped out and replaced with a new one without need for tools.

The Record Paper

The usual means of obtaining a printed record has been to use a carbon tape or a typewriter ribbon in conjunction with a record paper tape. The image of the counter numerals is imparted on the record tape when struck by a platen. The new design eliminates the carbon tape, spools, guides, and driving mechanism which have been necessary in the past. The elimination of these parts is made possible by having a single record tape specially coated to provide a surface which may be permanently marked by pressure against the coating. Fig. 3 illustrates the construction of the record paper. A layer of white paper is coated with black or other suitable dark color. This is then coated with a spongy whitish layer. Where the spongy layer is compressed or physically removed, the dark coating will show. The action of the platen on the paper when it strikes the counter figures causes the dark undercoating to show, duplicating on the record paper the figures of the counter at the instant of impact.

A further advantage of the specially coated record paper over the combination carbon or typewriter ribbon and record paper is that legible records are obtained over a greater temperature range. The new record tape can actually be used at −40 C (degrees centigrade) and up to 100 C. Heaters have been necessary in previous designs to keep the carbon tape usable at below freezing temperatures. Also, at temperatures over 60 C, the wax on carbon tape would melt and soon fail to produce a record.

The Impulse Accumulator and Governor

To obtain a maximum impulse count of 999 instead of the previous 299, a 3-figure resettable counter is used. Since between 1 and 2 seconds are required to print and then reset the printing counter, it was possible that one or two impulses from a watt-hour meter or other sensing device might be received during this interval. This could cause smeared printing if the counter figures were advancing at the same time that printing occurred, or an impulse would be lost completely if it should come in after printing occurred but before the counter had been completely reset. These difficulties have been overcome by a means of locking the counter drive shaft and a means of storing, then counting, any impulse which entered the meter while the counter driving gear was locked. To lock the counter drive shaft during printing and resetting, a locking cam is made to rotate to a position of interference with a pin in the driving gear of the counter. Since, with the locking cam in this position, the counter drive gear cannot rotate during reset, the gear meshed with it cannot rotate; see Fig. 2.

At this time the features of the accumulator will be discussed so that its relationship to the counter during and just after the completion of reset may be better understood. Fig. 4 shows the accumulator assembly, which consists of a shaft, and impulse motor-driven gear secured to the shaft; and a second gear loose on the shaft but held in position

---

Paper 58-146, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing December 6, 1957.

P. V. Terry is with the General Electric Company, Somersworth, N. H.

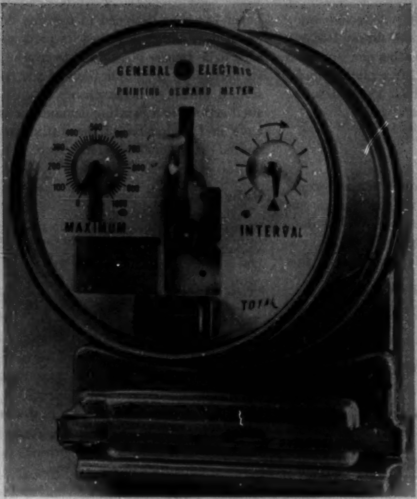


Fig. 1. New printing demand meter

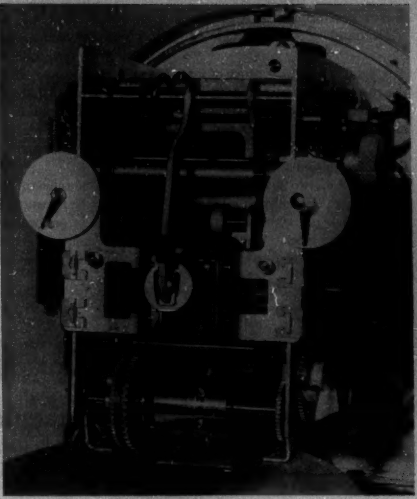


Fig. 2. Front view of new printing demand meter: faceplate down

relative to the first gear by means of a helical spring. A pin on each of the gears extends to overlap the other and prevent more than one revolution of relative motion between the two gears.

The two pins are biased together by the spring. The second gear rotates with the motor-driven gear and drives the printing counter except when the counter drive gear is locked. In this condition, the motor driven gear will rotate with each incoming impulse and cause the spring to wind up. At the completion of reset, the locking cam moves out of the way of the pin on the counter drive gear. If any impulses have been received during the reset period, the accumulator spring will unwind rotating the counter drive gear until the stop pins on the two accumulator gears meet. Any impulses received during reset are in this way registered on the printing counter.

It was found that the action on the spring unwinding was too fast to obtain proper life or registration of the impulses on the printing counter, and a way to slow this action was needed. Three standard methods of governing the speed of such an action were: 1. An air governor or fan, 2. an eddy current disk, and 3. a flyball or other centrifugal and frictional device.

All three of these methods were unsatisfactory because they would have to operate during each count or else use some sort of a differential to make the governor operate only if impulses are received during reset.

The method finally evolved consists of a ratchet wheel fastened to one of the accumulator gears and a rocking arm escapement mounted on the second accumulator gear. The inertial forces resulting as the weighted rocking arm escapement is caused to oscillate by the ratchet wheel keeps the speed of the spring-driven gear within bounds. The governor operates only when necessary after the completion of reset and no complicated differential gear trains are required. With this design, wear of the governor is not a problem.

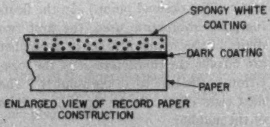


Fig. 3. Record paper construction

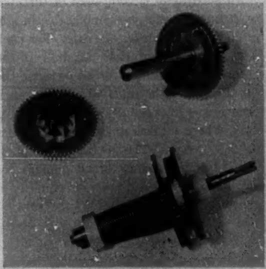


Fig. 4. Accumulator and governor

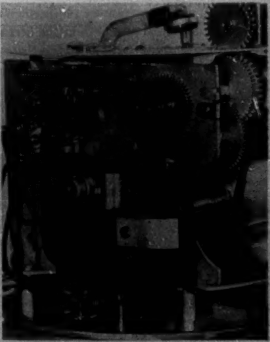


Fig. 5. Demand dial reset mechanism

Reset Mechanism of Running Demand

A detail which may prove of interest to the reader is the rack rod drive associated with the running demand, which is shown in Fig. 5. In an early design stage, two sets of bevel gears were employed to obtain a translation and rotation of the axis of rotation of two shafts. The same job is done more efficiently at a lower cost by using a rack rod which reciprocates in a nylon bearing block. The rack rod, as its name suggests, is a rod with annular rack teeth cut on it. A reciprocating pinion meshes with, and drives, one end of the rack rod. The other end transmits reciprocating motion to a pinion on the demand dial shaft which acts to reset to zero the running demand dial.

Conclusions

The value of a printing demand meter may be measured by the cost which it entails in performing its required function. To increase its design value, the new printing demand meter described in this paper has been designed to measure demand more accurately than its predecessors, to be smaller in size, and to cost less. In particular, the application of the described means of accomplishing the desired actions in a printing demand meter have in the opinion of the author advanced the art of demand metering.

Reference

1. 8th Annual Electrical Industry Forecast 1957–1970, Arthur J. Stegeman. *Electrical World*, New York, N. Y., vol. 148, Sept. 2, 1957, pp. 91–106.

The Use of a Digital Computer in a Generator Reserve Requirement Study

HOMER E. BROWN
MEMBER AIEE

A MODERATE-SIZED digital computer can be an effective tool in the calculation of the probability of simultaneous forced outages which are used in determining the generator reserve requirements of a utility system. The large amount of arithmetical work connected with an extensive study by the probability method makes the use of a digital computer imperative if the results are to be obtained in a reasonable length of time. This paper describes a program for the International Business Machines (IBM) *650* digital computer that has been written for this purpose.

The calculation of the probability of simultaneous forced outages used in determining generator reserve requirements has been discussed at length in the literature.[1–8] Reference 1 discusses the theory and reference 2 the application of the method of probability in determining generator reserve requirements as developed by Adler. In the following a program is described for the IBM *650* that was written specifically for this procedure of calculation. References 3 through 8 provide background for the subject but are not referred to directly.

The result of the probability calculations is a curve showing the relationship between forced outages in megawatts and the interval between such occurrences. The curve indicates the generator reserve that must be provided for the forced outage contingency to ensure the desired system reliability. Provisions for overhaul schedules and other factors must be determined in a separate calculation.

In Adler's method, a system is subdivided into subsystems in order to increase the accuracy of the results without an excessive amount of computation. Each subsystem is composed of turbines or generators that are comparable in size and have the same outage rate. For each subsystem, computations are made of the capacity lost through forced outages, the duration, and the interval between such occurrences for r simultaneous forced outages, where r takes on successive values $0,1,2,\ldots, r_{max}$. The maximum number, r_{max}, of simultaneous forced outages for small subsystems is usually set equal to n, the number of machines in the subsystem. In larger subsystems r_{max} is assigned a value large enough for the interval between such simultaneous occurrences to be several orders of magnitude greater than the desired reliability, decided upon by management. This facilitates the drawing of the forced outages in megawatts versus interval curves.

To obtain the probability, duration, etc., for the entire system it is necessary to combine the values for the various subsystems until the system has been reassembled. After each combining operation, the results must be grouped (condensed) before combining with another subsystem. This process is repeated over and over until the system is complete.

The final curve of forced outage in megawatts versus interval for the entire system permits management to decide upon the amount of reserve capacity necessary to protect the system against forced outages to the degree dictated by company policy.

Computations of forced outages are based upon the duration and interval between such outages for the different types of equipment as obtained from national averages and records of the system with which the author is connected. The results of these computations are the probability, duration, interval, frequency, capacity, and the standard deviation of the capacity for the various amounts of outage.

Discussion of Program

Because the decimal point swings through such a wide range, scaling is impossible. The calculations are therefore made in the floating decimal point mode. (In this mode the computer itself keeps track of the decimal point.) In the floating-point subroutine used, the first two digits in the representation of a number are 50 plus the characteristic of the logarithm, to the base 10, of the number. The next eight digits are the significant digits of the number.

Thus, 1229 is represented as 5312290000 and 0.001229 is represented by 471229-

---

Paper 58-134, recommended by the AIEE Computing Devices Committee and the System Engineering Committee of the Power Apparatus and Systems Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 21, 1957; made available for printing December 4, 1957.

Homer E. Brown is with the Commonwealth Edison Company, Chicago, Ill.

The author wishes to express his thanks to the members of the Analytical Engineering Department of the General Electric Company in Schenectady, N. Y., for their generosity in furnishing subroutines used in this program; and to H. A. Adler, J. P. Lynskey, and E. E. Ciesielski who explained their method of computation.

EXHIBIT NO. 1

```
                    INTERPRETATION  OF  DATA  CARDS  OF  PARTS  1 AND 2
   CARD
PRINT-OUT 59000076  51314400006  48112800006  51181700006  51340000006  51120000006  50600000006  52531900006
INTERPRETATION

   MEMORY  Mw.=31.44    p=.01128    σ=18.17      n=34.        r=12.        t=6.         T=531.9
  LOCATION
    0590

          59000076  52126040006  48112800006  51214700006  51130000006  51120000006  50600000006  52531900006

           Mw.=126.04   p=.01128    σ=21.47      n=13.        r=12         t=6.         T=531.9
```

Fig. 1. Print-out of two data cards for input to parts 1 and 2 of the problem

---

EXHIBIT NO. 2

```
 MEMORY               PARTIAL  PRINT  OUT  OF  RESULTS  OF  PART 1.
LOCATION  r OUTAGES        p            t            T            f           Mw.           σ
130000076                49679975346  51154676526  51227473726  48439611226

130700076 50100000006  49263759356  50435880116  51165256746  48605119046  51314400006  51181700006
NOTE:   FIRST  TWO  OF  TWELVE  LINES  ARE  SHOWN
```

EXHIBIT NO. 3

```
 MEMORY               PARTIAL  PRINT  OUT  OF  RESULTS  OF  PART 2.
LOCATION  r OUTAGES        p            t            T            f           Mw.           σ
 70000076                49862885696  51404538596  51468820606  48213301216

 70700076 50100000006  49127977146  50527734146  51412365946  48242503056  52126040006  51214700006
NOTE:   FIRST  TWO  OF  TWELVE  LINES  ARE  SHOWN
```

EXHIBIT NO. 4

```
 MEMORY               PARTIAL  PRINT  OUT  OF  RESULTS  OF  PART 3.
LOCATION  r OUTAGES        p            t            T            f           Mw.           σ
                         49586740996  51111893656  51190703656  48524373816

          50100000006  48870212996  50393483296  51452168956  48221156276  52126040006  51214700006
NOTE:   FIRST  TWO  OF  ONE  HUNDRED FORTY  FOUR  LINES  ARE  SHOWN
```

EXHIBIT NO. 5

```
 MEMORY               PARTIAL  PRINT  OUT  OF  RESULTS  OF  PART 4.
LOCATION  r OUTAGES        p            t            T            f           Mw.           σ
130000076                49586740996  51111893656  51190703656  48524373816

130700076 50100000006  49227594176  50393483286  51172888126  48578408746  51314400006  51181700006
NOTE:   FIRST  TWO  OF  THIRTY  SIX  LINES  ARE  SHOWN
```

EXHIBIT NO. 6

```
 MEMORY               PARTIAL  PRINT  OUT  OF  RESULTS  OF  PART 5.
LOCATION  r OUTAGES        p            t            T            f           Mw.           σ
 70000075                49586740996  51111893656  51190703656  48524373816

 70700076 50100000006  49227594176  50393483286  51172888126  48578408746  51314400006  51181700006
NOTE:   FIRST  TWO  OF  THIRTY  SIX  LINES  ARE  SHOWN
```

Fig. 2. Partial print-out of results of part 1, part 2, part 3, part 4, and part 5 of the program

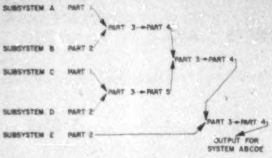
0000. Fig. 1 shows exhibit 1 which is a print-out of data cards and the interpretation of the floating-point representation. Fig. 2 shows exhibits 2 through 6 which are also in this form.

In the calculations of probability by Adler's method no special techniques are required. It is merely the volume of arithmetic that justified the use of a digital computer.

The calculations naturally divides itself into five parts and the program has been written in five subprograms in such a way that any part may be used separately or in combination with other parts, which greatly increases the flexibility of the program. The five parts are as follows:

1. Calculation of the probability of outage of one subsystem.

2. Calculation of the probability of a second subsystem.


SUBSYSTEM A PART 1
SUBSYSTEM B PART 2
SUBSYSTEM C PART 1
SUBSYSTEM D PART 2
SUBSYSTEM E PART 2
PART 3→PART 4
PART 3→PART 4
PART 3→PART 5
PART 3→PART 4
OUTPUT FOR SYSTEM ABCDE

NOTE: PART 3 MESHES RESULTS OF 1 & 2, OR 4 & 5, OR 1 & 5, OR 2 & 4.

Fig. 3. Flow diagram showing procedure of solving a problem involving five subsystems

3. Meshing the results of parts 1 and 2 to obtain the probability of outage considering both subsystems simultaneously.
4. Grouping (condensing) the results of part 3.
5. Grouping (condensing) the results of part 3 into a different memory location. (This is necessary so that the outputs of part 4 and 5 can become the input to part 3 in later phases of the study.) See Fig. 3.

In a probability study a set of machines of comparable size and type may be described by seven parameters, as follows:

1. p = the probability of forced outage
2. n = the number of machines
3. mw = average size, megawatts
4. t = the duration of forced outage
5. σ = standard deviation
6. r = maximum number of simultaneous outages to be considered
7. T = interval between outages

Note: T is not independent of t and p.

Details of the Program

From the national averages and system records t (the duration) and T (the interval between outages) for the particular equipment are obtained. This defines p, the probability of outage because $p = t/T$. Then, according to the theory developed, the probability, duration, interval between, etc., for r forced outages of n pieces of equipment to occur simultaneously may be computed.[1,2]

Parts 1 and 2 of the program perform the calculations given in Table I.

These values are stored in consecutive memory locations for the different values of r. When the results are punched out by a 7-word-per-card subroutine, each card contains the complete story for the particular outage of the subsystem. The listed results for two subsystems are shown in exhibits 2 and 3 in Fig. 2.

Part 3 performs the following calculations:

1. $p_{ab} = p_a \times p_b$
2. $t_{ab} = \dfrac{(t_a \times t_b)}{(t_a + t_b)}$
3. $T_{ab} = \dfrac{t_{ab}}{p_{ab}}$
4. $f_{ab} = \dfrac{1}{T_{ab}}$
5. $mw_{ab} = mw_a + mw_b$
6. $\sigma_{ab} = \sqrt{\sigma_a^2 + \sigma_b^2}$

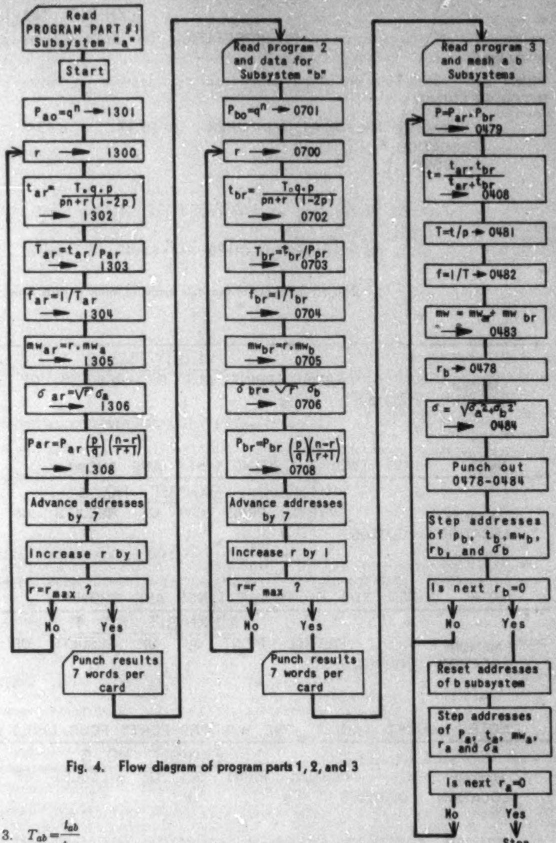
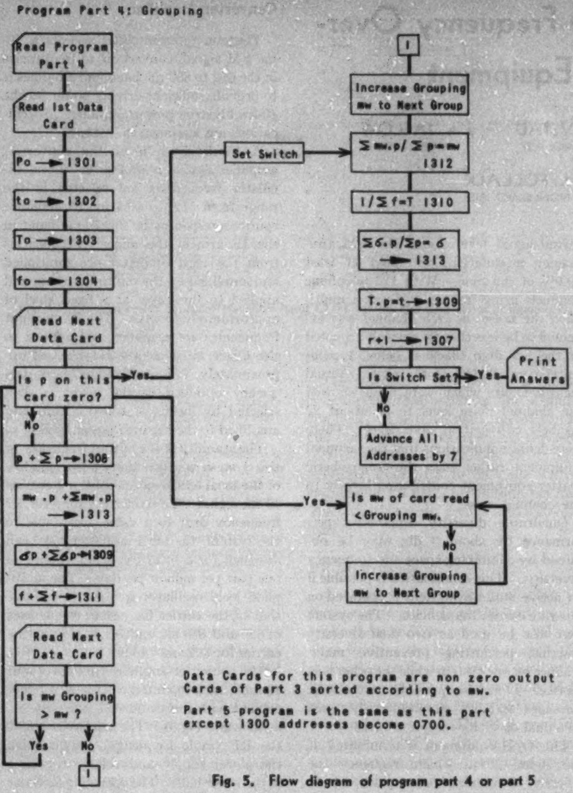

Read PROGRAM PART #1 Subsystem "a"
Start
P_ao=q^n → 1301
r → 1300
t_ar = T.q.p / (pn+r(1-2p)) → 1302
T_ar=t_ar/P_ar → 1303
f_ar=1/T_ar → 1304
mw_ar=r.mw_a → 1305
σ ar=√r σ_a → 1306
Par=P_ar (p/q)((n-r)/(r+1)) → 1308
Advance addresses by 7
Increase r by 1
r=r_max ?
No
Yes
Punch results 7 words per card
Read program 2 and data for Subsystem "b"
P_bo=q^n → 0701
r → 0700
t_br = T.q.p / (pn+r(1-2p)) → 0702
T_br=t_br/P_br → 0703
f_br=1/T_br → 0704
mw_br=r.mw_b → 0705
σ br= √r σ_b → 0706
P_br=P_br (p/q)((n-r)/(r+1)) → 0708
Advance addresses by 7
Increase r by 1
r=r max ?
No
Yes
Punch results 7 words per card
Read program 3 and mesh a b Subsystems
P=P_ar·P_br → 0479
t= t_ar·t_br / (t_ar+t_br) → 0408
T=t/p → 0481
f=1/T → 0482
mw = mw_ar + mw br → 0483
r_b → 0478
σ = √(σ_a²+σ_b²) → 0484
Punch out 0478-0484
Step addresses of p_b, t_b, mw_b, r_b, and σ_b
Is next r_b=0
No
Yes
Reset addresses of b subsystem
Step addresses of P_a, t_a, mw_a, r_a and σ_a
Is next r_a=0
No
Yes
Stop

Fig. 4. Flow diagram of program parts 1, 2, and 3

In these calculations, a runs the entire gamut from 0 to r_1 for the first subsystem for each value b from 0 to r_2 for the second subsystem; see Fig. 4 for details of the program.

The probability values obtained from this meshing operation form the p-matrix. The value p_{ab} is the result of meshing p_a of the first subsystem and p_b of the second system, and occupies the position in row a, column b of the matrix. The same is true of the results obtained for t, T, f, mw, and σ. The program is simultaneously computing the elements of six matrices. For two subsystems in which r is permitted to vary from 0 to 30, 900 elements would be produced for each of the six quantities. For each combination of a and b one card is punched which contains the complete information on this particular combination of outages (a outages in the first subsystem, and b in the second). These cards become the input cards for parts 4 and 5, and are therefore punched out in load card format, i.e., a plus punch is emitted in column 1. These output cards are sorted according to magnitude of mw by a card sorter before entering parts 4 or 5 as data. The results are grouped (condensed) into blocks according to the magnitude of mw. The

Vol.
Pt. 1
1958


Program Part 4: Grouping
Read Program Part 4
Read 1st Data Card
Po → 1301
to → 1302
To → 1303
fo → 1304
Read Next Data Card
Is p on this card zero?
Yes
No
p + Σ p → 1306
mw . p + Σmw . p → 1313
σp + Σσp → 1309
f + Σf → 1311
Read Next Data Card
Is mw Grouping > mw ?
Yes
No
1
Set Switch
1
Increase Grouping mw to Next Group
Σ mw.p/ Σ p = mw 1312
1/Σ f = T 1310
Σσ.p/Σp = σ → 1313
T.p = t → 1309
r+1 → 1307
Is Switch Set?
Yes
Print Answers
No
Advance All Addresses by 7
Yes
Is mw of card read < Grouping mw.
No
Increase Grouping mw to Next Group

Data Cards for this program are non zero output Cards of Part 3 sorted according to mw.

Part 5 program is the same as this part except 1300 addresses become 0700.

Fig. 5. Flow diagram of program part 4 or part 5

size of the blocks is controlled by the investigator. The calculations performed by parts 4 and 5 follow:

1. $p_{a'} = \Sigma p_{ab}$
2. $mw_{a'} = (\Sigma p_{ab} \times mw_{ab})/\Sigma p_{ab}$
3. $\sigma_{a'} = (\Sigma p_{ab} \times \sigma_{ab})/\Sigma p_{ab}$
4. $f_{a'} = \Sigma f_{ab}$
5. $T_{a'} = 1/f_{a'}$
6. $t_{a'} = T_{a'} \times p_{a'}$

In all these summations, $(a'\text{-}1)\ mw_g < mw_{ab} < a' \times mw_g$, where mw_g is the magnitude of the grouping *mw*'s. The output cards of part 3 are read and included in the summations until the next value of *mw*'s excedes the limit for the particular block. The calculations for this group are then completed and the program advances to the next group. This continues until the data are exhausted; see Fig. 5 for details of the program.

Parts 1 and 2 are identical except for the memory block assigned to them for storage of their results. The same is true of parts 4 and 5. Parts 1 and 4 use the same memory block (1300–1899). Parts 2 and 5 use the same memory (700–1299). Part 3 meshes the results it finds in the locations 0700–1299 and 1300–1899 regardless of their origin (original calculations of parts 1 and 2 or grouping operation of parts 4 and 5).

The flow diagram illustrates the method of using this program for a system with five subsystems; see Fig. 3.

Table I

No.	Calculation	Explanation
1	r takes on successive values 0,1,2	
2	$p_r = \frac{n!}{r!(n-r)!}(1-p)^{n-r} \times p^r$	Probability that r of n machines with probability of outage rate p would fail simultaneously.
3	$t_r = \frac{T \times (1-p)p}{p(n-r) + r(1-p)}$	Duration of such simulteanous occurrences.
4	$T_r = \frac{t_r}{p_r}$	Interval between such occurrences.
5	$f_r = \frac{1}{T_r}$	Frequency.
6	$mw_r = r \times mw$	Capacity of equipment involved, megawatts.
7	$\sigma_r = \sqrt{r} \times \sigma$	Standard deviation of capacity, measure of uncertainty.

Recapitulation

If the theory of probability is accepted for determining the amount of generator reserve requirement, this program can obtain the results quickly with a great economy of engineering time. In an extensive study 100 man-weeks of hand calculation were produced in 8 hours on the *IBM 650*. The program has been used several times with excellent results.

The use of the computer relieves the engineer of the drudgery of such a study so that he may spend more time on the technical aspects of the work. The division of the program into five parts greatly increases the versatility of the program.

References

1. A NEW APPROACH TO PROBABILITY PROBLEMS IN ELECTRICAL ENGINEERING, H. A. Adler, K. W. Miller. *AIEE Transactions*, vol. 65, Oct. 1946, pp. 630–32.

2. DETERMINATION OF RESERVE GENERATING CAPABILITY, H. Halperin, H. A. Adler. *Ibid.*, vol. 77, pt. III, 1958 (*Paper no. 58-126*).

3. GENERATING RESERVE CAPACITY DETERMINED BY THE PROBABILITY METHOD, Giuseppe Calabrese. *Ibid.*, vol. 66, 1947, pp. 1439–50.

4. DETERMINATION OF RESERVE CAPACITY BY THE PROBABILITY METHOD, G. Calabrese. *Ibid.*, vol. 69, pt. II. 1950, pp. 1681–89.

5. CALCULATING PROBABILITY OF GENERATING CAPACITY OUTAGES, W. J. Lyman. *Ibid.*, vol. 66, 1947, pp. 1471–77.

6. PROBABILITY METHODS APPLIED TO GENERATING CAPACITY PROBLEMS OF A COMBINED HYDRO AND STEAM SYSTEM, E. S. Loane, C. W. Watchorn. *Ibid.*, pp. 1645–57.

7. A CONVENIENT METHOD OF DETERMINING GENERATOR RESERVE, Howard P. Seelye. *Ibid.*, vol. 68, pt. II, 1949, pp. 1317–20.

8. ELEMENTS OF SYSTEM CAPACITY REQUIREMENTS, C. W. Watchorn. *Ibid.*, vol. 70, pt. II, 1951, pp. 1163–85.

Wide-Band Ultrahigh-Frequency Over-the-Horizon Equipment

R. A. FELSENHELD
NONMEMBER AIEE

H. HAVSTAD
NONMEMBER AIEE

J. L. JATLOW
MEMBER AIEE

D. J. LeVINE
ASSOCIATE MEMBER AIEE

L. POLLACK
NONMEMBER AIEE

Synopsis: Wide-band O/H (over-the-horizon) radio equipment in the 680- to 900-mc (megacycle) band suitable for use in toll-quality multichannel telephone or TV (television) circuits is described. The system comprises 60-foot parabolic antennas fed with dual polarization horns, 10-watt drivers, 10-kw amplifiers using 6-cavity klystrons, and receivers which permit dual or quadruple diversity by combining the received signals at the IF (intermediate frequency). The over-all 1-db (decibel) bandwidth is 15 mc, and the time delay distortion characteristics are suitable for interconnection with existing toll-quality radio links.

THE BROAD-BAND tropospheric communication system described can transmit between 120 and 600 telephone channels with trunk circuit quality, or a TV program, up to at least 200 miles. The equipment also interconnects with common carrier telephone networks.

For the 185-mile link between Florida and Cuba, with a path loss of 197 db, this equipment, employing space diversity, was expected to provide a mean carrier-to-noise ratio of about 42 db and a received-signal level above the FM (frequency modulation) threshold at least 99.9% of the time. With 120 telephone channels using frequency division multiplex, the noise on each channel was expected to be less than 26 decibels adjusted at the −9 dbm (decibels below 1 milliwatt) point 99.9% of the time. Actual measurements which were made showed the channel noise level to be about 22 decibels adjusted at zero level. There were definite indications that the terminal equipment rather than the tropospheric scatter equipment contributed largely to the channel noise.

Quadruple diversity, improving performance by about 3 db, may be obtained by combining space and frequency diversity. This combination is possible if an active stand-by system is operated on separate carrier frequencies. The system may also be used as two dual-diversity systems, permitting preventive maintenance on one system while the other is in service. It is also possible to transmit messages and TV simultaneously, each on a dual-diversity reception basis.

This O/H equipment is terminated at the 70-mc IF at which frequency the diversity combining takes place. It is thus possible to connect to either line-of-sight or O/H repeaters at IF instead of base-band frequency. Table I lists the specifications for the system, and Fig. 1 shows its major elements and the switching facility provided.

The first application of this broad-band equipment is the microwave communication system between Miami and Havana jointly sponsored by The American Telephone and Telegraph Company and the International Telephone and Telegraph Company. The system includes two *TD-2* line-of-sight hops in Florida between Miami and Florida City where it connects with the O/H hop to Guanabo, Cuba. A *TD-2* line-of-sight hop completes the system from Guanabo to Havana. This system supplements the present cable system between Florida and Havana to provide many more message channels. It also provides a TV channel between the United States and Cuba.

Converter-Amplifier

The converter-amplifier accepts a 70-mc FM signal, converts it to frequencies in the 692 to 880-mc band and amplifies it to provide sufficient driving power for the 10-kw klystron power amplifier. Specifications are as shown in Table II.

A block diagram of the converter-amplifier is shown in Fig. 2. Local oscillator frequencies are selected in the range from 51.8 to 59.4 mc to minimize spurious responses in the 70-mc band in the IF arm of the mixer. Frequencies from the local oscillator are multiplied and amplified by the multiplier chain and applied to the mixer at a fixed level of approximately 4 watts. The 70-mc input frequencies are amplified and applied to the mixer at an adjustable level of approximately 1/2 watt. The beat frequency obtained in the mixer is then selected by means of tuned circuits and amplified to the required power level.

The stability of the output frequency of this system is a function of the stability of the local oscillator and of the incoming 70-mc signal. To reduce the total carrier frequency drift to a value acceptable to the system, the local oscillator has been designed for a total frequency change of one part per million per day. The multiplied local oscillator frequency is above that of the carrier for output frequencies of 840 and 880 mc and below that of the carrier for 692 and 740 mc.

The converter-amplifier consists of four 19-inch, rack-mounted chassis in a 7-foot cabinet with front and rear access doors, as shown in Fig. 3. The top half contains the RF (radio frequency) chassis, with the power supply and voltage regulator in the lower half. The operating facilities on the front which are accessible without interrupting the interlocks are: the metering panel, tuning of local oscillator and multipliers, and tuning of mixer amplifier. The last two are available through doors in the protective dust cover. Power control and circuit breakers, and a line voltage meter are available on the front panel. A ventilating fan, with replaceable filter, is mounted on the inside of the rear door. The rear door is interlocked with the high-voltage controlling relays and when opened the high voltage is turned off.

RF CHASSIS

The RF chassis, Fig. 4, in addition to providing a mounting surface for the detachable IF power amplifier and RF multiplier subchassis, also contains the mixer-amplifier chain, metering facilities, monitoring circuits, fan and manifold, interlocks, and alarm circuitry.

Paper 58-179, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 26, 1957; made available for printing December 2, 1957.

R. A. FELSENHELD, H. HAVSTAD, J. L. JATLOW, D. J. LEVINE, and L. POLLACK are with the Federal Telecommunication Laboratories, Nutley, N. J.

The successful development of this equipment is due to the teamwork and co-operation of a large number of people. The authors are grateful to the following engineers for their contributions: A. G. Kandoian, for the initial planning and over-all supervision; W. Sichak, for over-all technical supervision; J. Gulack for the system layout; W. Glomb, H. E. Lott, and J. Simkovich, for the converter-amplifier; L. Gray, H. Goldman, and W. Jacobus, for the power amplifier; R. T. Adams, A. T. Brown, and R. Bresk, for the receiver; and L. Juhas, for the feed horn. The requirements and equipment specifications for the system were set up by a joint committee of representatives of the Bell Telephone Laboratories, the Long Lines Department of the American Telephone and Telegraph Company, the Western Electric Company, the International Telephone and Telegraph Corporation, and the Federal Telecommunication Laboratories. Special acknowledgment is due to the following members of the Committee: K. P. Stiles and A. A. Bottani of the Long Lines Department, American Telephone and Telegraph Company; and H. E. Curtis, N. F. Schlaack, and H. A. Wells of the Bell Telephone Laboratories.

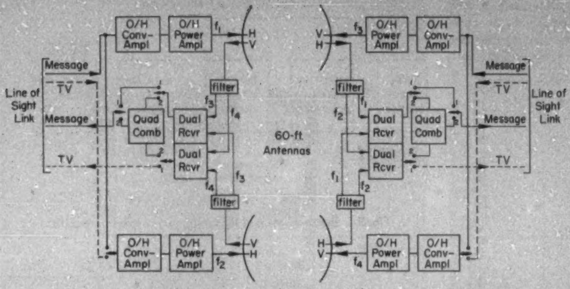


Fig. 1. Block diagram of complete wide-band O/H system showing major components

Table I. System Specifications

Features	Requirements
Transmitter power	10 kw.
Antennas	60-foot parabolic reflectors (two at each end) with dual polarization feed horns.
Diversity	Quadruple (space and frequency); may be used as two dual diversity systems; IF (70 mc) combining used.
Service	System input and output 70-mc FM signal; capable of handling at least 120 telephone channels; may also be used simultaneously for TV transmission.
Frequency band	680–900 mc.
Transmission lines	*WR-1150* wave guide for transmitting line, pressurized; $3^1/_8$-inch Styroflex cable for receiving line, pressurized.
Bandwidth	15 mc, 1 db down; 19 mc, 3 db down: with suitable terminal equipment, 6 mc, 1 db down video to video.
Envelope delay distortion	Suitable for multihop, multichannel telephone or wide-band television transmission; see text.
Noise loading distortion products	More than 50 db below signal level.
Safety features	All equipment fully interlocked and protected; key interlocks used in power amplifier.
Alarm and metering circuits	All equipment fully equipped with alarm circuits and contacts for external alarms; metering and test points incorporated for alignment and maintenance.
Environmental	Continuous operation in tropical climates.

Three types of metering and monitoring circuits are available.

1. Two panel meters to monitor certain points which indicate satisfactory operation of the equipment.
2. Test points, on a jack panel accessible without turning the equipment off, with the proper test equipment to determine transmission characteristics of the equipment.
3. Each vacuum tube is connected to a test jack for measuring the cathode current by means of one of the panel meters.

The ventilation system is located on the rear of the chassis and consists of a 115-cubic-foot-per-minute blower with a distribution manifold and individual air ducts leading to each tube to be cooled. Failure of the blower is indicated by a relay and alarm buzzer activated by an air-flow switch, as well as by a lamp located on top of the cabinet.

Two devices are used for protection against exposure to high voltage. The first device employs interlock switches in the front dust covers and rear cabinet access door. The second employs gravity-operated short-circuiting bars, operated by removal of the front dust covers and the rear access door.

STABLE OSCILLATOR

The oscillator consists of a stable, crystal-controlled oscillator and a buffer amplifier. The tubes used are one *12AT7* and one *404A*. The crystal is temperature-controlled and operates in its fifth overtone from 51.8 to 59.4 mc with a stability of 1.0 parts per million per day. The output is about 2 volts into a 90-ohm load.

RF MULTIPLIER

A chain of four tubes multiplies the frequency received from the local oscillator and amplifies the power to a level sufficient to drive the high-level mixer. The multiplication factors are 12 for output frequencies of 622 to 670 mc and 16 for output frequencies of 910 to 950 mc. The tubes used are one *2C51/396A*, one *418A*, and two *2C39B*'s. The relative power levels and frequency distributions along the multiplier chain are shown in Fig. 5. One volt across the 90-ohm input impedance is required for a 3-watt output.

The coupling circuits used between the first three tubes are conventional, lumped-constant, tuned circuits with a radial transmission-line cavity terminating the fourth tube. An adjustable coupling loop in this cavity couples the power to the input of the mixer.

The envelope temperature is kept below approximately 100 degrees centigrade by means of air cooling on all tubes except the *2C51*, and the total power dissipation in any tube is below 50% of the rated dissipation for the tube.

IF POWER AMPLIFIER

The IF amplifier is a broad-band 6-tube 5-stage unit which amplifies the 70-mc signal. It will deliver an output power adjustable between 0.1 to 1.5 watts

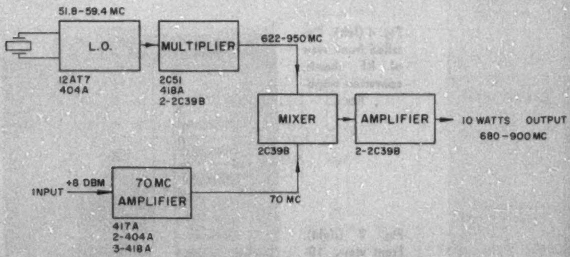


Fig. 2. Block diagram of converter-amplifier

Table II. Specifications for Converter-Amplifier

Features	Requirements
Input	
Power	+8 dbm minimum
Frequency	70 mc
Bandwidth	20 mc
Impedance	75 ohms
Output	
Power	10 watts
Frequency	680–900 mc
Bandwidth	20 mc at 1/4-db points
Impedance	50 ohms
Envelope delay	<4 millimicroseconds for ±4 mc

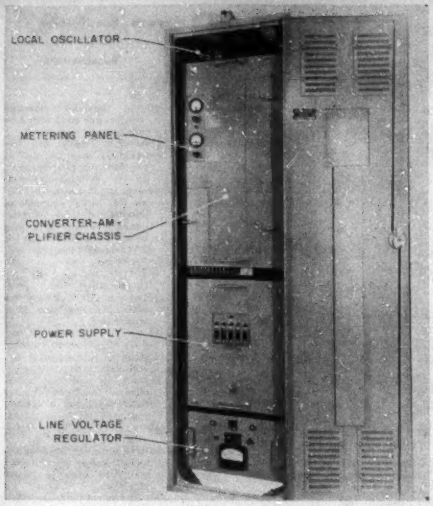

LOCAL OSCILLATOR
METERING PANEL
CONVERTER-AMPLIFIER CHASSIS
POWER SUPPLY
LINE VOLTAGE REGULATOR

Fig. 3. Front view of converter-amplifier

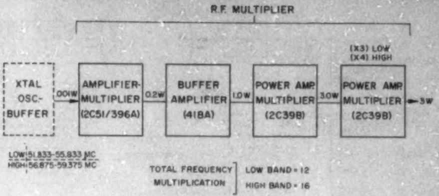

R.F. MULTIPLIER
XTAL OSC-BUFFER
.001W
AMPLIFIER-MULTIPLIER (2C51/396A)
0.2W
BUFFER AMPLIFIER (418A)
1.0W
POWER AMP. MULTIPLIER (2C39B)
3.0W
(X3) LOW
(X4) HIGH
POWER AMP. MULTIPLIER (2C39B)
3W
LOW 51.833-55.833 MC
HIGH 56.875-59.375 MC
TOTAL FREQUENCY MULTIPLICATION
LOW BAND = 12
HIGH BAND = 16

Fig. 5. Block diagram of RF multiplier, converter-amplifier

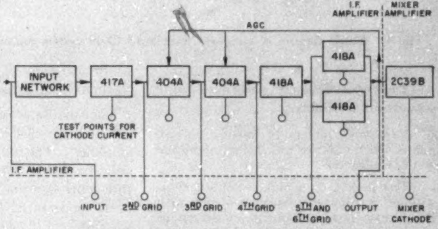

I.F. AMPLIFIER
MIXER AMPLIFIER
AGC
INPUT NETWORK
417A
404A
404A
418A
418A
418A
2C39B
TEST POINTS FOR CATHODE CURRENT
I.F. AMPLIFIER
INPUT
2ND GRID
3RD GRID
4TH GRID
5TH AND 6TH GRID
OUTPUT
MIXER CATHODE

Fig. 6. Block diagram of IF amplifier, converter-amplifier

with an input signal of 7 milliwatts across 75 ohms. The tubes used are one *417A* grounded-grid input stage, two *404A*'s and one *418A* interstage, and two *418A*'s in a push-pull output stage; see Fig. 6.

The input coupling network is provided with a matching adjustment consisting of bias control of the input tube and a single-tuned broad-band cathode inductance. A matching adjustment is necessary because the input to this amplifier may be fed by a relatively long transmission line, and also to minimize distortion due to reflections.

Most interstage coupling networks are adjustable equivalent T transformers with a passband amplitude response of 0.1 db over a 20-mc band centered at 70 mc. Exceptions are the coupling between the fourth and fifth stages, and the output transformer where fixed coupled transformers are used.

Three methods are used to minimize variations in amplifier gain resulting from tube aging and supply voltage fluctuations. These consist of: 1. regulation of the filament power supplies by use of a

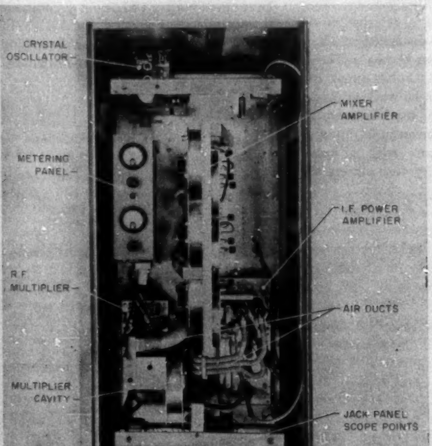

CRYSTAL OSCILLATOR
METERING PANEL
R.F. MULTIPLIER
MULTIPLIER CAVITY
MIXER AMPLIFIER
I.F. POWER AMPLIFIER
AIR DUCTS
JACK PANEL SCOPE POINTS

Fig. 4 (left). Detailed front view of RF chassis, converter-amplifier

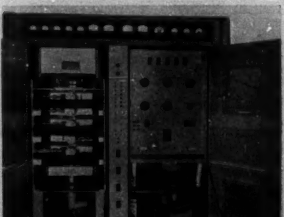


Fig. 7 (right). Front view, 10-kw power amplifier

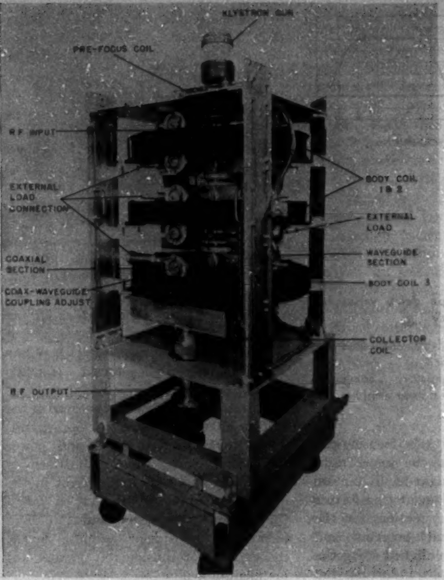

KLYSTRON GUN
PRE-FOCUS COIL
R F INPUT
EXTERNAL LOAD CONNECTION
COAXIAL SECTION
COAX-WAVEGUIDE COUPLING ADJUST
R.F. OUTPUT
BODY COIL 1 & 2
EXTERNAL LOAD
WAVEGUIDE SECTION
BODY COIL 3
COLLECTOR COIL

Fig. 8 (left). Klystron assembly and magnetic structure mounted on dolly

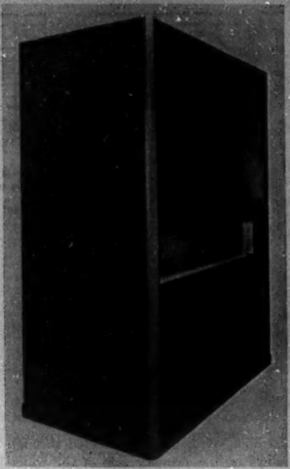


Fig. 10 (right). Heat exchanger: coolant to air

12-volt filament supply and appropriate dropping resistors, 2. positive grid bias and large cathode resistors, and 3. *AGC* feedback from IF output to the grids of the second and third stages.

MIXER-AMPLIFIER

The mixer-amplifier consists of three *2C39B*'s in double-tuned radial transmission line cavity-coupling networks. The required bandwidth is obtained by using stagger-damped interstage coupling and single-tuned broad-band output coupling, giving an over-all 3-peaked response having a peak-to-valley ratio of approximately 1/4 db over a 20-mc bandwidth.

The mixer-amplifier, Fig. 4, consists of a mixer, two ultrahigh-frequency amplifiers, and radial transmission line cavities. Each tube has one cathode cavity and one plate cavity. Broad bandwidth is obtained by adjusting the coupling between the plate cavity of one stage and the cathode cavity of the following stage. The local oscillator power is coupled to the cathode of the mixer tube by means of a radial cavity, and the IF power is introduced through a broad-band lumped-constant network. The plate cavity of the mixer is loop-coupled to the cathode cavity of the first amplifier, and proper coupling is obtained by the relative values of two tuning capacitors in the amplifier cathode cavity.

A similar coupling arrangement is used between the first and second amplifiers. The second amplifier plate circuit is a single tuned cavity, loaded to the required Q by the 50-ohm load on the output loop. Capacitive-tuning plungers, inserted parallel to the axis, tune each cavity.

POWER SUPPLY

The power supply is unregulated and supplies all direct and alternating voltages. Voltage regulation of ±1% is obtained from an automatic, line-voltage regulator mounted below the power supply.

A 60-second time delay between the application of filament power and plate power provides time for filament warm-up.

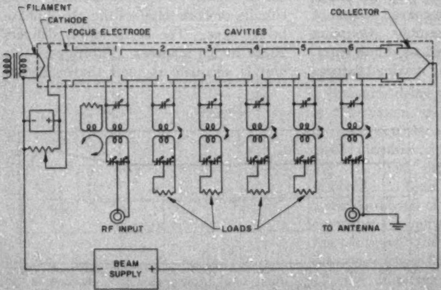

FILAMENT
CATHODE
FOCUS ELECTRODE
CAVITIES
COLLECTOR
1
2
3
4
5
6
RF INPUT
LOADS
TO ANTENNA
BEAM SUPPLY
-
+

Fig. 9 (left). Equivalent circuit of 6-cavity klystron

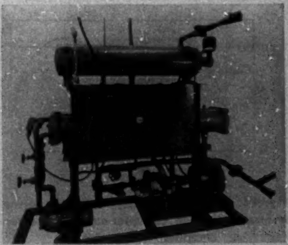


Fig. 11 (right). Heat exchanger: coolant to water

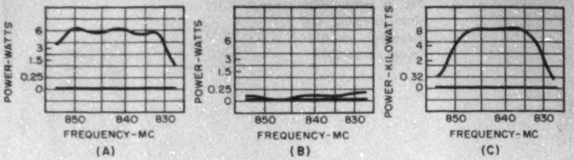


Fig. 12. **Amplitude response versus frequency**

A—Converter-amplifier
B—Reflected from klystron input
C—Ten-kw amplifier driven by converter-amplifier

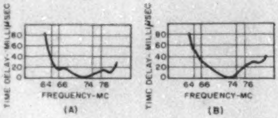


Fig. 13 (left). **Envelope delay versus frequency**

A—Modulator, demodulator, and converter-amplifier
B—Modulator, demodulator, converter-amplifier, and 10-kw power amplifier

All high voltages are connected through relays on the primary side of the transformers which are operated by the interlocks on the rear cabinet door and front dust covers.

Power Amplifier

The converter-amplifier output at approximately 10 watts is amplified to 10-kilowatt level by the power amplifier. The required 30 db of gain is obtained with a 6-cavity klystron (type *X631*, by Eitel-McCullough, Inc.). This klystron and a set of tuned circuits covers the band from 680 to 900 mc.

Amplifier

Since the velocity modulation action at the gaps and the density modulation in the drift spaces are frequency-sensitive mechanisms, the klystron is inherently a narrow-band device. However, by utilizing broad-band tuned circuits at the klystron gaps, bandwidths of 3% to 4% of center frequency can be achieved.

The amplifier assembly is shown in Fig. 7. The klystron assembly is on the left and the controls for the focusing-coil power supplies are on the upper right panel. The lower right-hand section houses the klystron filament transformer and the semiconductor rectifier for the focus supplies. Circuit breakers and overload indicators are aligned along the center panel, and meters indicating the klystron parameters are in the top panel.

The klystron assembly, Fig. 8, shows in greater detail the double-tuned circuits surrounding each klystron gap. The circuits across the first five klystron gaps consist of coaxial lines coupled to rectangular wave-guide cavities. A wave-guide section surrounds the klystron gap. To handle the high power level, the output circuit is formed by two wave-guide cavities coupled together with an iris located in the common wall between the guide sections. A lumped-circuit equivalent of these tuned circuits is shown in Fig. 9.

The desired bandwidth is achieved by loading each circuit with an external resistor. In the input circuit, the wave-guide section is loaded and the input signal is coupled to the coaxial section of the double-tuned assembly. Coupling and loading are adjusted to obtain maximum power transfer and maximal flatness of response (equal primary and secondary Q).[1] The interstage assemblies are loaded on one side only (the secondary coaxial section) to achieve the maximum impedance across the klystron gap for the given bandwidth.

The output primary tuned circuit is loaded by the klystron beam impedance; the secondary is loaded by the antenna impedance coupled through a capacitive probe. The loading resistors used in the first three tuned circuits are air cooled. The fourth and fifth circuit resistors are $1\frac{5}{8}$-inch coaxial-line water-filled loads.

To simplify the alignment of the amplifier, the six tuned circuits are synchronously aligned and centered about the carrier frequency. Each circuit is initially adjusted for transitional coupling with the klystron unenergized. A sweep signal generator is coupled to one circuit at a time and the response observed at the test point incorporated in the circuit. Then, under normal drive and power input conditions, a slight readjustment of the circuit parameters will result in the desired response. It should be noted that increased efficiency and gain may be obtained by "stagger damping" but at the expense of considerably greater adjustment difficulties.[2]

Power Supplies

The primary power to all filaments, to the klystron bombarder supply, and to the electromagnetic focusing supplies is regulated by an automatic motor-driven induction regulator to assure maximum tube life and minimum variation in klystron-operating conditions.

All high-voltage magnetic components are nonflammable-oil-filled. The bombarder supply and beam supply are unitized assemblies, i.e., the magnetic components are contained in one oil-filled tank.

The bombarder supply utilizes Xenon-filled rectifiers, whereas the beam supply utilizes high-vacuum diodes to minimize the possibility of rectifier arcback. The circuits of both supplies are the 3-phase bridge type. The beam supply delivers 20 kv at 2.5 amperes. Filtering of this unit is adequate to limit the incidental

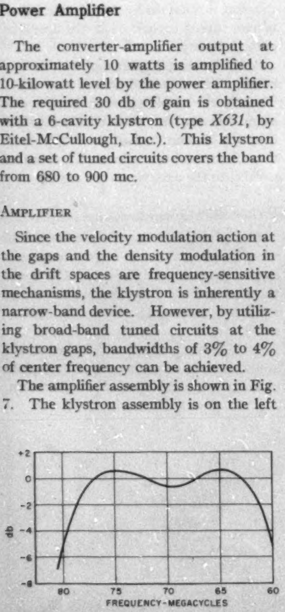


Fig. 14. **Over-all amplitude response versus frequency through converter-amplifier, power amplifier, and receiver**

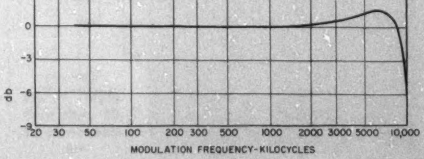


Fig. 15. **Over-all amplitude video-to-video amplitude response, including wide-band modulator and demodulator**

Fig. 16. Sixty-foot parabolic antennas at Guanabo, Cuba

FM to less than 100 cycles per second rms. Continuous variation of the beam voltage is possible at the amplifier control panel, over the range from 10.5 to 20 kv at the nominal line voltage ±10%.

Controls

Control and overload monitoring circuits are arranged to provide proper sequential application of the various voltages and to protect the equipment against progressive damage resulting from component failures or overloads.

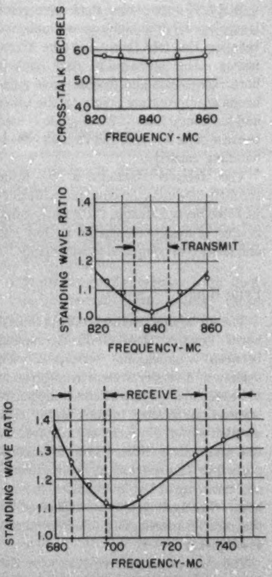


Fig. 17. Standing wave ratio and crosstalk of a typical dual-polarization feed horn

All a-c circuits are monitored and switched by magnetic-trip circuit breakers with an appropriate tripping time versus current characteristic which will accommodate starting current surges but also allow rapid tripping under large overload conditions.

High-speed d-c relays monitor the following parameters: beam current, drift tube current, fifth and sixth cavity voltage, excessive output transmission line mismatch, and low RF output. All of these relays except the low RF output will trip the beam power supply circuit breaker in less than 50 milliseconds. Pilot-light indications and an alarm calls attention to a fault. In addition to electrical interlocks, the necessary mechanical interlocks such as liquid coolant flow, air flow, and high-voltage grounding switches are incorporated.

Cooling

Since the collector and drift tubes require liquid cooling, a closed system (coolant to air or coolant to water unit) is provided for each amplifier. Where an adequate supply of well water is available, the latter type of cooler is preferred. The air-cooled unit is shown in Fig. 10 and the water-cooled unit in Fig. 11. The coolant, a mixture of ethylene glycol-distilled water (60 to 40 parts per volume) with an acidic inhibitor, is also used as the dissipative element in the klystron fourth and fifth cavity loads.

Performance

Over-all transmission response of the converter-amplifier, plus power amplifier is shown in Fig. 12; the oscillogram shown was taken at 840 mc. Note that the 3-db bandwidth is 19 mc, and the 1.0-db points approximately 15 mc. Saturation power is 10.5 kw at the upper end of the band (880 mc) and it falls to 8 kw at the low end of the band (690 mc).

Gain at the level 0.5 db below saturation power is 32 db at the high-frequency end of the band and 28 db at the low end for the bandwidths shown. For smaller bandwidths, the gain can be increased to as much as 48 to 50 db before instability problems arise.

Envelope delay versus input frequency of the system, including the wide-band terminal equipment, converter-amplifier, power amplifier, and receiver is shown in Fig. 13. The envelope delay curve is resolved into three components to obtain a standard basis of comparison; the slope component is found by drawing a straight line between the frequency points of interest, such as 66 mc and 74 mc of Fig. 13; the difference in time delay noted between the two points is the slope. The change in envelope delay remaining after subtracting the slope component is the parabolic component, provided ripple is negligible.

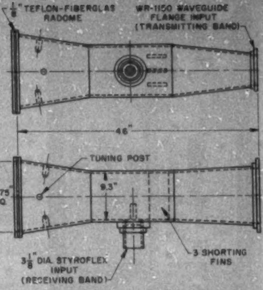

$\frac{1}{8}$" TEFLON-FIBERGLAS RADOME
WR-1150 WAVEGUIDE FLANGE INPUT (TRANSMITTING BAND)
46"
TUNING POST
13.75" SQ.
9.3"
$3\frac{1}{8}$" DIA. STYROFLEX INPUT (RECEIVING BAND)
3 SHORTING FINS

Fig. 18. Outline drawing of dual-polarization feed horn

If appreciable ripple is present, this must be averaged about the slope component and subtracted from the total envelope delay; the remainder is the parabolic component. Thus, the sum of the three components at each frequency should equal the recorded curve. The components of time delay are as follows:

	Time Delay, Millimicroseconds	
Component	**At ±4 Mc**	**At ±6 Mc**
Parabola	15	22
Slope	25	30
Ripple	±2	±2

Over-all response from IF input to the converter-amplifier to IF output at the receiver is shown in Fig. 14, and the video-to-video response from the transmitting terminal through the radio system and the receiving terminal is shown in Fig. 15. Intermodulation distortion is 50 db down, as measured through the system with a 500-kc bandwidth noise source producing an rms deviation of 0.4 mc.

Fig. 19. View of dual-polarization feed horn

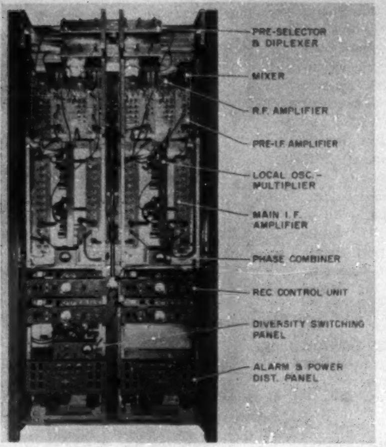

PRE-SELECTOR & DIPLEXER
MIXER
R.F. AMPLIFIER
PRE-I.F. AMPLIFIER
LOCAL OSC.-MULTIPLIER
MAIN I.F. AMPLIFIER
PHASE COMBINER
REC CONTROL UNIT
DIVERSITY SWITCHING PANEL
ALARM & POWER DIST. PANEL

Fig. 20 (left). Front view of quadruple diversity receiver

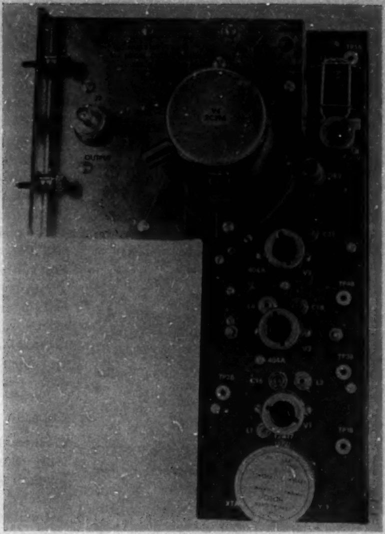


Fig. 24 (right). Crystal-controlled local oscillator

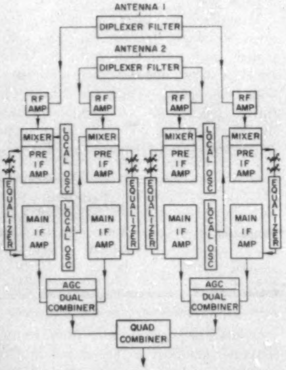

ANTENNA 1
DIPLEXER FILTER
ANTENNA 2
DIPLEXER FILTER
RF AMP
RF AMP
RF AMP
RF AMP
MIXER
MIXER
MIXER
MIXER
PRE IF AMP
PRE IF AMP
PRE IF AMP
PRE IF AMP
LOCAL OSC
LOCAL OSC
LOCAL OSC
LOCAL OSC
EQUALIZER
EQUALIZER
EQUALIZER
EQUALIZER
MAIN IF AMP
MAIN IF AMP
MAIN IF AMP
MAIN IF AMP
AGC
DUAL COMBINER
AGC
DUAL COMBINER
QUAD COMBINER

Fig. 21 (left). Block diagram of major receiver components

Antenna System

The antenna system consists of two 60-foot paraboloidal reflectors at each site, fed by dual polarized horns, each of which is connected to one transmitter and two receivers. The transmission line connecting the transmitters to the antennas is *WR-1150* wave guide, and the receiver connection is via $3^1/_8$-inch 50-ohm Styroflex cables.

REFLECTORS

The reflectors and feed-horn support structures are shown in Fig. 16.

TRANSMISSION LINES

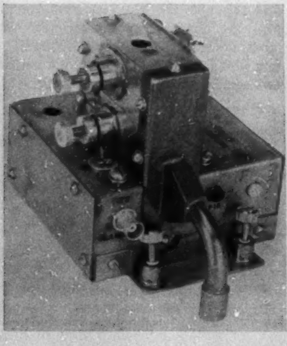
WR-1150 copper-clad steel wave guide is employed for the high-power connection between the feed horn and the transmitter. The wave-guide runs include both rigid bends and flexible wave-guide bends at appropriate points on the tower and horizontal lines. Wave guide is used because of its low loss, and its high power-handling capacity.

The receiver connection is made through $3^1/_8$-inch, 50-ohm polyethylene-jacketed Styroflex cable. This large cable is used primarily because of its low loss and lower cost compared to wave guide.

FEED HORN

The use of dual, orthogonally polarized horns (greater than 50-db decoupling between polarizations) permits a wide variety of diversity system arrangements to be employed. In particular, the polarizations are selected in such a way that no high-power diplexers are required in the transmitter lines. This is accomplished by transmitting at one frequency and receiving at two frequencies, in each horn. Received signals are separated by the receiver-preselecting filters arranged as a diplexer.

The dual polarized feed horns for the 60-foot paraboloidal reflectors were designed to direct a signal at the reflector edge at a level of approximately −10 to



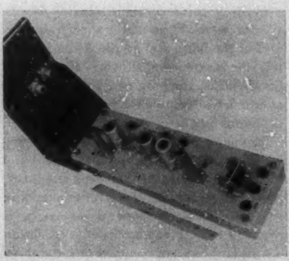


Fig. 23. Microstrip mixer and 70-mc preamplifier

Fig. 22 (left). Low noise RF amplifier

−12 db relative to the signal directed toward the reflector vertex.[3] Typical data are given in Fig. 17. Since the horns were designed for a specific application, no particular effort was directed toward broad-band development.

An outline drawing of the horn is shown in Fig. 18. Short capacitive posts were employed to watch the horn at the individual transmitter frequency bands. This was necessary to obtain the required low VSWR (voltage standing wave ratio), less than 1.08, over the transmitter band. The receiver connections were matched by proper selection of probe length and diameter and by the fin position relative to the probes.

The horns have been satisfactorily tested with 10-kw continuous-wave power fed into the wave-guide connection. The transmission lines and horns will operate with approximately 1/2-psi (pound-per-square-inch) (gage) internal dry air pressure, and have been tested with over 2-psi (gage) internal pressure.

The horn aperture barrier is a gasketed 1/8-inch-thick teflon-fiberglass sheet clamped to the aperture flange. Fig. 19 shows a completed horn, ready for installation. The horn weighs approximately 150 pounds. The maximum dimensions are approximately 14 by 14 by 46 inches.

The antenna system performance may be described as follows: The free-space gain at mid-band is 41 db, not including feed line loss, and the half-power beamwidth is 1.4 degrees, mid-band.

Diversity Receiver

The quadruple diversity receiver set comprises two dual-diversity reciever sets operating at different frequencies in the range 680 to 900 mc with means for combining at the 70-mc IF frequency into one 70-mc output.

Fig. 20 shows the receiver with dust covers removed. Two 9-foot racks (joined together at the center) hold the various 19-inch subchassis panels. Available components such as IF amplifiers, equalizers, and receiver control units, are used wherever possible.[4] Preselectors, mixers, local oscillator-multipliers, combiners, diversity switching panel, power distribution and alarm panels are designed specifically for this system.

Fig. 21 is a simplified block diagram. Preselectors are provided to separate the two frequencies from each antenna line for each of the four receiving channels. Each signal passes through an RF amplifier and mixes with the output of a crystal-controlled local oscillator multiplier to produce a 70-mc IF. The pre-IF amplifiers, equalizers, attenuators, and main IF amplifiers deliver the proper level for phase combining. There are three phase combiners, one for each dual-diversity channel, and one to combine the dual-diversity outputs for quadruple combining. To meet the over-all delay distortion and amplitude response requirements, each major component of the receiver was so designed that the 0.1-db bandwidth is about 20 mc.

Diplexer Filter

The diplexer is composed of two 3-cavity maximally flat amplitude response filters, centered at 692 and 740 mc respectively, (or 840 and 880 mc) coupled to a T-junction. The branch legs to each filter are a quarter wavelength at the center frequency of the other filter so that short circuits at the filter inputs are reflected as open circuits at the diplexer junctions. The correct leg is a matched line (SWR under 1.1). The filter specifications are:

Response, Db Bandwidth	Megacycles	Insertion Loss	Decibels
0.1	18.5	center frequency	1/2
1.0	30	other received frequency	30
3.0	38	transmitter frequencies	50
40	175		

RF Amplifier

A single *416B* tube is used in a grounded grid RF amplifier stage. Fig. 22 is a top view. The input SWR is less than 1.3 over a 20-mc bandwidth. The plate circuit has two resonant transmission lines provided with adjustable coupling and an output impedance of 50 ohms. Gains of 14 db and noise figures under 8 db are obtained.

Mixer Preamplifier

Microstrip construction is used in the balanced mixer.[5] The crystals (types *1N21B* and *1N21BR*) are each matched to a 50-ohm line by a short-circuited stub of proper length and position. Decoupling between the signal input and the local oscillator arm is greater than 20 db. The signal input SWR is under 1.3 for a 20-mc bandwidth. The mixer is attached directly to the pre-IF amplifier, Fig. 23, and feeds the cathode of the first of three cascaded stages.

Main IF Amplifier

A wide-band IF amplifier, eight tubes in cascade, is used to provide 60-db gain at 70-mc center frequency.[4] The 3-db bandwidth is 32 mc, and the 0.1-db bandwidth is 20 mc.

Local Oscillator: Multiplier

The crystal-controlled local oscillator, Fig. 24, uses a *12AT7* twin triode in a cathode coupled circuit at frequencies in the 50-mc range. The output triode section is tuned to twice the crystal frequency. Two *404*-type tubes follow in cascade, the first as a 100-mc amplifier and the second as 300-mc tripler. This drives the grounded grid output (*2C39*) tube either as a doubler or tripler in the frequency range 620 to 950 mc.

Phase Combiner, 70-Mc

Diversity reception is made possible by locking two 70-mc IF signals and adding them in phase. The phase detector supplies a d-c control voltage to the local oscillator, which in turn adjusts their frequencies to produce phase lock. Except for two semiconductor diodes, the combiner uses only passive circuit elements. Part of the combined output voltage is used for automatic gain control and controls the main IF amplifiers. The 70-mc outputs of the dual combiner connect to the diversity switch panel.

Diversity Switching Panel

Outputs from the dual combiners are switched by coaxial relays to separate jacks or to a third combiner for quadruple operation. Automatic gain control voltages are simultaneously switched for either type of service. Fig. 21 shows the system diagram for a quadruple receiver.

Power Distribution and Alarm Panels

Each dual-diversity receiver is separately powered by external lines from a commercial power supply. Incoming power is controlled by master switches, and circuit breakers and fuses. The alarm circuitry monitors the blower operation, the RF level, and the IF output; and 3-phase lock conditions actuate an audible alarm whenever their operation is abnormal.

References

1. Design Charts for Tuned Transformers. M. J. Hellstrom. *Electronics*, New York, N. Y., vol. 29, no. 11, Nov. 1956, p. 182.

2. Vacuum Tube Amplifiers, George E. Valley, Jr., Henry Wallman, editors. McGraw-Hill Book Company, Inc., New York, N. Y., MIT Radiation Laboratory Series, vol. 18, 1948, p. 221.

3. Dual-Mode Feed Horn for Microwave Multiplexing, D. J. LeVine, W. Sichak. *Electronics*, vol. 27, no. 9, Sept. 1954, p. 162.

4. TD-2 Microwave Radio Relay System, A. A. Roetken, K. D. Smith, R. W. Friis. *Bell System Technical Journal*, New York, N. Y., vol. XXX, Oct. 1951, pp. 1041–77.

5. Microstrip—A New Transmission Technique for the Kilomegacycle Range, D. D. Grieg, H. F. Engelmann. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Dec. 1952, pp. 1644–63.

The Miami-Havana Radio System and Its Integration Into the Telephone Networks

K. P. STILES
MEMBER AIEE

F. G. HOLLINS
ASSOCIATE MEMBER AIEE

E. T. FRUHNER
NONMEMBER AIEE

W. D. SIDDALL
NONMEMBER AIEE

A MIAMI-HAVANA radio system, consisting of the wide-band over-the-horizon system discussed in another paper,[1] together with *TD-2* microwave links at each end, has recently been placed in service. This paper discusses the over-all radio system, its method of operation, and some of the test results obtained during its initial line-up.

Over-All System Description

The Miami-Havana radio system, developed to provide both telephone and television facilities, is divided into three sections. Starting south from Miami, Fla., two links of *TD-2* microwave equipment (totaling 37 miles) are used to connect the Miami terminal equipment to the over-the-horizon equipment at the Florida City radio station. From Florida City to Guanabo, Cuba, the 185-mile jump is made by means of the over-the-horizon equipment previously described; see Fig. 1. Between Guanabo and Havana a 14-mile *TD-2* microwave link is used to connect the over-the-horizon equipment to the terminal equipment. The standard *TD-2* FM (frequency modulation) signal is passed to and from the over-the-horizon equipment at a 70-mc (megacycle) intermediate frequency. The FM terminals for both telephone and television services are located at Miami and Havana.

Two separate radio systems are provided in each direction. At each end the *TD-2* links are operated in the common carrier band of 3,700 to 4,200 mc. At Florida City the over-the-horizon transmitters operate on 840 and 880 mc, and at Guanabo frequencies of 692 and 740 mc are used. Since these frequencies are in the ultrahigh-frequency television range they were co-ordinated with existing and proposed television station assignments in southern Florida to minimize chances of interference. Each over-the-horizon transmitter frequency is received on a space diversity basis, and arrangements are available to permit reception on a quadruple-diversity basis when required. A block schematic is shown in Fig. 2.

Since the radio station at Florida City is constructed in the Everglades, southeast of Florida City, an area subject to flooding under unusually high tide conditions, it was necessary to fill in the area used for construction of the building and the antenna assemblies. All construction is of a type designed to withstand winds of hurricane force. Fig. 3 is a view toward the horizon from the Florida City station. At Guanabo the station is located on the crest of a hill some 200 feet above the terrain immediately in front of it and has an unobstructed path to the horizon. Fig. 4 is a view toward the horizon from the Guanabo station.

The Florida City–Guanabo radio system provides two radio space diversity channels ordinarily used on a regular and protection basis. With automatic switching in the *TD-2* links, and the over-the-horizon channels fed in parallel from the *TD-2* channels selected by the automatic switching equipment, the over-all system is designed to provide as nearly continuous service as feasible. At Miami and Havana *L*-type carrier terminals provide connections to the land line systems of the two countries. In Miami this carrier terminal equipment is located in the telephone building where all other message circuits are terminated. In Havana the terminal equipment is located in the Radio Corporation of Cuba offices and connected by cable to the switchboards located in the Cuban Telephone Company office about a mile distant.

For television service, television operation centers are located at the Telephone Company office in Miami and the Havana office of the Radio Corporation of Cuba, and connections to television networks or stations are made from these points. When television service is required, use is made of the protection facilities and the telephone system is operated without benefit of the protection switching arrangements. Fig. 5 is a view of the front of the Florida City station.

Transmission Characteristics

During the initial tests, recordings were made of the incoming RF (radio frequency) signal level both to compare transmission on the two individual radio systems and to verify the path loss of 197 db (decibels), between isotropic antennas, previously determined. The two radio systems showed the same average fading characteristics although the instantaneous input levels were usually different. The 197-db path loss was verified, but recordings will be continued for some time to determine what daily and seasonal characteristics the path may have.

The basic transmission characteristics of the beyond-the-horizon path vary continuously because of selective fading resulting from multipath transmission. This amplitude and phase distortion is reflected in instability in the recovered base-band signal. Since the side bands furthest removed from the carrier of the FM wave are most affected in amplitude and phase relative to the carrier by selective fading, it is only to be expected that this instability would increase with base-band frequency. Relatively stable transmission was observed up to about 1 mc, with increasing instability to the top of the video band.

However, the higher frequency components of a working monochrome television signal are quite small, so that occasional loss or distortion of these components will cause picture distortions which are usually apparent only to a trained observer. This characteristic makes it possible to provide a good television picture on the over-the-horizon system. However, momentary loss of horizontal or vertical synchronization is immediately evident, and it is necessary that these signal components be relatively undistorted. The relatively stable characteristic of the low frequencies in the over-the-horizon system meet this requirement. In contrast to television, multiplex telephone transmission distributes the energy of individual components across the band much more uniformly. It would not therefore be possible on this system to occupy fully the same frequency band employed for television with a similar band of telephone channels. Telephone channels are thus limited to a base band of about 1 mc.

Paper 58-154, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 18, 1957; made available for printing November 12, 1957.

K. P. STILES, F. G. HOLLINS, and E. T. FRUHNER are with the American Telephone and Telegraph Company, Long Lines Department, New York, N. Y.; and W. D. SIDDALL is with the Radio Corporation of Cuba, Havana, Cuba.

The construction of the over-all Miami-Havana radio system was the result of co-operation by several companies, which include the Radio Corporation of Cuba, the Federal Telecommunication Laboratories, the Long Lines Department of the American Telephone and Telegraph Company, the International Telephone and Telegraph Company, and the Bell Telephone Laboratories.

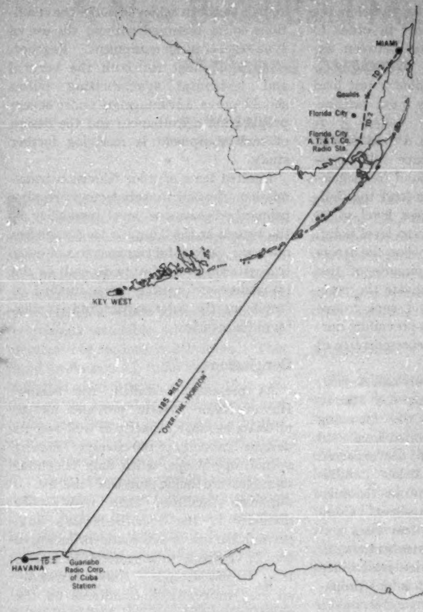

MIAMI
TD 2
Goulds
Florida City
TD 2
Florida City
A.T.&T. Co.
Radio Sta.
KEY WEST
185 MILES
"OVER-THE-HORIZON"
HAVANA
TD 2
Guanabo
Radio Corp.
of Cuba
Station

Fig. 3. View toward horizon at Florida City

Fig. 4. View toward horizon at Guanabo, Cuba

Fig. 1 (left). Layout map of Florida-Cuba over-the-horizon radio system

This over-the-horizon system presently employs FM with a deviation of ±1 mc. The base band initially employed for the telephone circuits is 300 to 550 kc. The base band for television service is the nominal bandwidth of the television signal. The telephone and television signals are pre-emphasized with the use of the corresponding pre-emphasis networks presently in use on type *TD-2* facilities. Tropospheric selective-fading characteristics require additional limiting for FM demodulation to that required by line-of-sight systems such as the type *TD-2* system. This over-the-horizon system is therefore provided with 50 db of limiting which provides reasonable stability of base-band output level.

As the telephone or television signal components are confined within a band of approximately 9 mc around the carrier, the received band is correspondingly limited to this by a filter to minimize noise and improve the system fade margin. Similarly, adjustment of the transmitter bandwidth to 11 mc permits improved operation of the over-the-horizon transmitters. At the Florida terminal, where transmitting frequencies of 840 and 880 mc are used, output powers of 8 to 10 kw were obtained with a d-c input of about 45 kw when tuned to provide maximum bandwidth. When the amplifier cavities were tapered to give an output band-

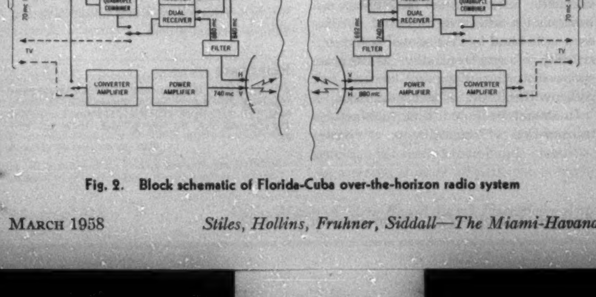

RADIO CORP. OF CUBA STATION
GUANABO, CUBA
TRANSMITS 692 AND 740 MEGACYCLES
VERTICALLY POLARIZED
A.T.& T. CO. STATION
FLORIDA CITY, FLA.
TRANSMITS 840 AND 880 MEGACYCLES,
HORIZONTALLY POLARIZED
185 MILES
LINE OF SIGHT
MICROWAVE
(1 HOP TD-2
TO HAVANA)
LINE OF SIGHT
MICROWAVE
(2 HOPS TD 2
TO MIAMI)
CONVERTER AMPLIFIER
POWER AMPLIFIER
692 mc
840 mc
V
H
FILTER
MESSAGE
QUADRUPLE COMBINER
DUAL RECEIVER
DUAL RECEIVER
840 mc
880 mc
740 mc
692 mc
70 mc IF
TV
740 mc
880 mc

Fig. 2. Block schematic of Florida-Cuba over-the-horizon radio system

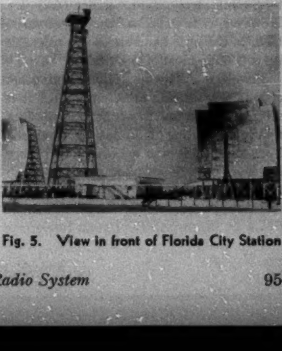


Fig. 5. View in front of Florida City Station

width of about 11 mc, the efficiency increased sufficiently to permit the same output power with a beam voltage of 19 kv as compared with 20 kv for the wider band output. At the Cuban terminal, some increase in output power was obtained by similar methods. These increases in output power further improved the fading margin.

The median RF signal-to-noise ratio is about 47 db with the adjustments indicated in the foregoing. Variations occur in the received signal level, and path loss tests on this path indicate that long-term fades in excess of 20 db below the median value may occur about 2% of the time. Even for average signals of 20 db below the median value, the signal-to-noise ratio should be quite adequate. For example, with the median RF signal-to-noise ratio of 47 db, the base-band video signal-to-noise ratio will be 48 db on the basis of peak-to-peak signal to rms noise.

The FM deviation of any system is, of course, a compromise between fluctuation noise at low deviations and distortions or instability at high deviations. Although a deviation of ±1 mc is presently employed as an optimum, future equipment developments may allow greater deviations. The over-the-horizon transmitting and receiving equipments will not be a limiting factor since these equipments are designed to provide for bandwidths considerably wider than those now in use.

Accordingly, the use of pre-emphasized FM with a nominal deviation of ±1 mc and with limiting of 50 db, as well as the conventional restriction of the equipment bandwidth to that required for the signal products, provides satisfactory transmission of high-capacity multiplexed telephone signals or monochrome television signals.

Telephone Service

The telephone circuits are presently subject to level variations of less than ±1.0 db over protracted periods. As indicated previously, these circuits occupy base-band frequencies under 550 kv. However, it is expected that level variations in the base band should be tolerable up to 1.0 mc most of the time.

Envelope delay distortions and other crosstalk-producing factors are such that strong single-frequency tones will produce relatively strong modulation products. For example, it was necessary to reduce the level of a 64-kc pilot tone by 10 db below the normal level employed on type *TD-2* systems to a level 24 db below the nominal telephone channel in order to minimize cross-modulation between circuits separated by 64 kc. With this modification, the system handled a test load of 180 telephone circuits quite satisfactorily.

Tests were made with a load of 180 telephone circuits, or three full supergroups, employing base-band frequencies up to 750 kc. With this load the individual circuits had a noise level of 22 decibels adjusted at the zero level point. This is about 10 db better than the design objective for telephone circuits of this length. These results indicate the practicability of operating four supergroups (240 circuits) if conditions prevailing during the initial period are representative of future conditions.

During the interval in July and August, 1957, available for testing the Miami-Havana radio systems, the incoming median RF signal levels varied from −10 to +20 db with respect to the expected median of about 50 db below 1 milliwatt. Under these conditions the telephone transmissions are excellent. Computations based on path loss tests and available data on fading rates and magnitudes indicate that a service availability may be expected of 99.99% from a propagation standpoint.

Television Service

Although tropospheric propagation has been employed elsewhere for a limited number of telephone circuits, this project represents the first attempt at television transmission. No information was available regarding envelope delay distortion, response variations, and other characteristics. In view of this, the over-the-horizon facilities were designed to have a sufficiently wide bandwidth to permit optimum operating conditions for television without restrictions or limitations from the facilities themselves.

The principal problem affecting television transmission is the stabilization of the synchronizing components. The transmission features of this system discussed in the foregoing were effective in providing the needed stabilization. For example, tests of non-pre-emphasized transmission with a deviation of ±4 mc on a wide-band basis disclosed severe instability of synchronization as well as severe distortion of picture content, together with substantial level variations.

It should be noted that satisfactory transmission of monochrome television signals has been achieved under the conditions so far observed without the use of sync-regenerating equipment. Regenerating equipment for both the vertical and horizontal synchronizing pulses should prove advantageous under severe propagation disturbances and the design of such equipment is receiving further study.

Limited tests of color television transmission showed unsatisfactory results, primarily because of level instability of the system at the 3.6-mc color carrier frequency. Successful transmission of color transmission in the future depends on the possibility of devising some method of stabilizing the color carrier and its side-band frequencies.

Conclusions

As presently installed, the Miami-Havana radio system provides for an ultimate message capacity of well over its designed capacity of 120 circuits. During normal operations, when only telephone messages are being handled, the use of duplicate facilities, each with space diversity in the over-the-horizon link, permits the use of automatic switching in the *TD-2* links, with restoration in less than 75 milliseconds. Under extremely adverse transmission conditions on the over-the-horizon section, quadruple diversity, a combination of both space and frequency diversity, may be used.

In conclusion, it appears that the use of over-the-horizon tropospheric scatter links of the type described will provide reliable radio systems for telephone or television service. Where terrain conditions are such that single hops of the order of 200 miles are required, systems of this type offer advantages over other transmission methods. It appears that where intermediate repeaters can be provided and maintained, line-of-sight transmission is better in propagation stability, has comparatively less multipath fading, provides greater availability of frequency assignments, and is lower in cost. The over-the horizon systems are expected to be particularly advantageous for connecting islands, for crossing isolated regions with rugged intermediate terrain, and in areas where frequency congestion in the radio frequency bands between about 500 and 2,000 mc is not severe.

Reference

1. WIDE-BAND ULTRAHIGH-FREQUENCY OVER-THE-HORIZON EQUIPMENT, R. A. Felsenheld, H. Havstad, J. L. Jatlow, D. J. LeVino, L. Pollack. *AIEE Transactions*, vol. 77, pt. I, Mar. 1958, pp. 86–93.

A New Broad-Band Microwave Antenna System

R. W. FRIIS
NONMEMBER AIEE

A. S. MAY
NONMEMBER AIEE

RADIO RELAY in the Bell Telephone System now provides over 13 million miles of 2-way message circuits, approximately one fourth of all the long-distance telephone circuits in this country. It also provides about 62,000 miles of broad-band radio channels which make network TV (television) available to over 500 TV broadcasting stations in 350 cities. This rapid expansion of radio relay facilities indicated some time ago that by 1955 portions of the central transcontinental and other backbone routes would be loaded to their capacity of six broad-band channels in each direction of transmission. Continued expansion would require the development of new radio relay systems in the higher frequency bands assigned by the Federal Communications Commission for common-carrier use.

A new antenna called the "horn-reflector" antenna[1,2] and its associated circular waveguide were the first items developed for these new higher frequency radio relay systems. The horn-reflector antenna and its circular waveguide are designed for simultaneous operation in the common-carrier bands of 3,700 to 4,200 mc (megacycles) (*TD-2*), 5,925 to 6,425 mc (*TH*), and 10,700 to 11,700 mc (*TJ*). The letters indicate the abbreviated Bell System code for the respective radio relay system, the latter two of which are under development. The 6,000-mc *TH* system will be a long-haul system with eight broad-band channels in each direction of transmission, each channel capable of carrying about 2,000 message circuits, or a monochrome or National Television Systems Committee color television program. The 11,000-mc *TJ* system will be a short-haul system, each broad-band channel capable of carrying up to 240 message circuits or a monochrome or NTSC color television program.

Fig. 1 shows an installation of the horn-reflector antenna at Terrell, Tex., on the 4,000-mc *TD-2* radio relay system from Atlanta, Ga., to Dallas, Tex. The installation of this broad-band antenna and waveguide promises considerable savings when 6,000-mc long-haul and 11,000-mc short-haul systems are added to this and other backbone routes in the future.

Application

Fig. 2 illustrates that the horn-reflector antenna is basically an electromagnetic horn capped by a sector of a paraboloidal-shaped reflector. The apex of the horn coincides with the focal point of the paraboloidal reflector. The latter converts the spherical wavefront emanating from the feed horn to a plane or uniphase wavefront at the antenna aperture, thus ensuring maximum gain and minimum side lobe radiation.

The feed horn at the focal point of the antenna is designed with the proper hyperbolic contour to match a 2.812-inch-ID (inside diameter) circular waveguide with the frequency range 3,700 to 11,700 mc. A circular copper waveguide with a 2.812-inch ID has three important advantages over rectangular waveguides: 1. the loss for the dominant TE_{11}^{0} mode transmission at 4,000 mc is about 0.42 db (decibels) per 100 feet, approximately one half the loss of a 2.290×1.145 rectangular copper waveguide; 2. two signals polarized at 90 degrees to each other may be transmitted through circular waveguide, and by careful control of manufacturing processes a cross-polarization discrimination greater than 40 db per $12^{1}/_{2}$-foot section can be achieved; and 3. it is possible to transmit simultaneously dual-polarized signals in the 4,000-, 6,000-, and 11,000-mc common-carrier bands.

Propagation measurements over a typical radio link employing circular waveguide and horn-reflector antennas indicate that the net cross-polarization discrimination between vertically and horizontally polarized signals is greater than 20 db. By alternately polarizing adjacent channels of the 6,000-mc *TH* system this 20 db of discrimination or selectivity may be used to reduce the spacing between broad-band channels. As a result, a more efficient use is made of the frequency spectrum and a greater number of message circuits or television channels can be allocated in the 5,925- to 6,425-mc common-carrier band.

Because of their broad-band characteristics, only four horn-reflector antennas will be required at each repeater station on main backbone routes to handle all three systems at 4,000, 6,000, and 11,000 mc. As shown in Fig. 1, a transmitting and receiving antenna will be used on main backbone routes for each of the two directions of transmission. On spur or light routes, where no more than three or four broad-band channels are utilized in any one common-carrier band, one antenna may be used for both transmitting and receiving in one direction, thereby reducing to two the number of antennas per repeater station.

Fig. 3 shows schematically a horn-reflector antenna and circular waveguide system in which a finline[3] polarization separation network *A* may be used to separate the vertically polarized signals from the horizontally polarized signals in the 4,000-, 6,000-, and 11,000-mc systems. The signals then pass through system separation networks[4] *B*, *C*, *D* which separate the signals for each of the three common-carrier bands. Constant resistance networks[5] similar to those developed for the *TD-2* system may then be used to separate the various channels in a particular common-carrier band.

The tuning section illustrated in Fig. 3 is used at the bottom of each waveguide run to improve the cross-polarization discrimination. It is a 12-inch section of waveguide purposely deformed to have a cross-polarization discrimination (XPD) of 25 db, a value approximating the net XPD of the antenna and a long waveguide run. As one of the final steps in the alignment of an antenna waveguide system, the tuning section is rotated to improve the XPD to a minimum value of 30 db.

Transmission Characteristics

GAIN

The gain of the antenna with respect to an isotropic radiator or point source is shown in Table I. Also the efficiency of the antenna or the ratio of the measured gain to a computed area gain for the 64.5-square-foot aperture is indicated in decibels down from area gain, and in per cent.

HORIZONTAL AND VERTICAL DIRECTIVITY

Fig. 4 shows a typical radiation pattern on a 360-degree polar diagram for the horizontal directivity (azimuth) at 3,740

Paper 58-185, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted July 26, 1956; made available for printing November 20, 1957.

R. W. FRIIS and A. S. MAY are with the Bell Telephone Laboratories, Inc., North Andover, Mass.

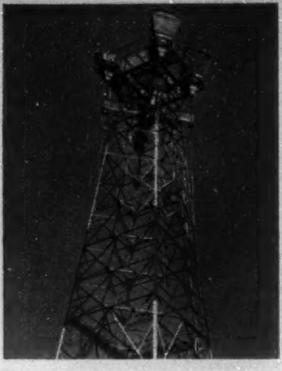


Fig. 1. Horn-reflector antenna installation at Terrell, Tex.

mc. The width of the beam between three decibel points, at 3,740, 6,325, and 10,960 mc is shown in Table II. At the higher frequencies the radiation patterns are similar to those shown in Fig. 4, except that even more of the energy is concentrated in the desired direction of propagation.

The vertical directivity (elevation) for both polarizations is also listed in Table II for the same frequencies.

The front-to-back ratio of the antenna, or its ability to discriminate against signals coming from the back (180-degree response), may be taken directly from the horizontal directivity patterns.

New antenna-measuring equipment has been developed which uses radar pulse techniques to achieve a higher order of accuracy in measurements of radiation patterns and crosstalk between antennas. In the past, measurements have been limited at times because of interfering reflections from nearby trees, buildings, and surrounding terrain. Magnetrons were developed for spot frequencies in the three common-carrier bands, and by their use the aforementioned extraneous reflections may be separated from the direct signal on a time basis, and thus a true plot of the radiation pattern and crosstalk obtained. However, the measured front-to-back ratios in Table II or the crosstalk characteristics in Table III may be reduced at a typical antenna installation because of ground return or reflections from the surrounding terrain.

CROSSTALK

The crosstalk or coupling between similar antennas in operating position at the same level on a tower is of particular importance when a 2-frequency allocation plan is used for the radio system. In this type of plan, different frequencies are assigned for transmitting and receiving, but the receiving antenna pointing to the west receives signals on the same frequency assignments as the receiving antenna pointing to the east. This is true also for the two transmitting antenna. Table III shows the measured back-to-back and side-to-side crosstalk of two antennas spaced on 20-foot centers. A pulsed test signal at a known level is sent into a transmitting antenna and the response or crosstalk is measured for different orientations of the receiving antenna.

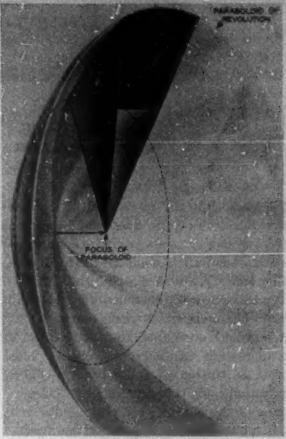


Fig. 2. Derivation of horn-reflector antenna

Table I. Antenna Gain Characteristic

Frequency, Mc	Polarization	Midband Gain, Decibels	Efficiency, Db Down from Area Gain	Efficiency, Per Cent
3,950	Vertical	39.6	1.6	69
	Horizontal	39.4	1.8	66
6,175	Vertical	43.2	1.8	66
	Horizontal	43.0	2.0	63
11,200	Vertical	48.0	2.2	60
	Horizontal	47.4	2.8	52

Table II. Antenna Directivity Characteristics

Frequency, Mc	Polarization	Front-to-Back Ratio, Db	Horizontal Directivity, Azimuth, Degrees	Vertical Directivity, Elevation, Degrees
3,740	Vertical	71	2.5	2.0
	Horizontal	77	1.6	2.13
6,325	Vertical	71	1.5	1.25
	Horizontal	71	1.25	1.38
10,960	Vertical	78	1.0	0.75
	Horizontal	71	0.8	0.88

CROSS-POLARIZATION DISCRIMINATION

The XPD of an antenna is a measure of the ability of the antenna to discriminate between two received signals at the same frequency, that are polarized 90 degrees with respect to each other. The maximum XPD of the horn-reflector antenna is shown in Table IV.

Fig. 5 shows the XPD of the main lobe of the horn-reflector antenna. Over the rest of the radiation pattern the responses to vertically and horizontally polarized signals are approximately equal. The utilization of cross-polarization to discriminate against interfering signals is therefore not feasible except within the directivity of the main lobe. Fig. 5 also illustrates how rapidly the XPD deteriorates if the antennas are misaligned or if their alignment is influenced by wind velocity, tower twist, or sway.

Table III. Antenna Crosstalk Characteristics

Frequency, Mc	Polarization	Crosstalk: Antennas Back-to-Back, Db	Crosstalk: Antennas Side-to-Side, Db
3,740	Vertical	140	81
	Horizontal	122	89
6,325	Vertical	140	120
	Horizontal	127	122
10,960	Vertical	139	94
	Horizontal	140	112

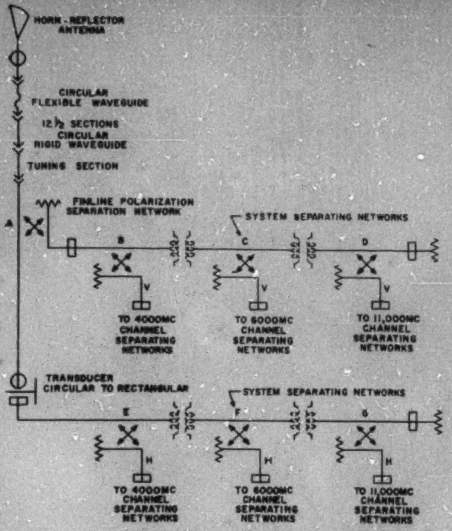

HORN-REFLECTOR ANTENNA
CIRCULAR FLEXIBLE WAVEGUIDE
12½ SECTIONS CIRCULAR RIGID WAVEGUIDE
TUNING SECTION
FINLINE POLARIZATION SEPARATION NETWORK
A
SYSTEM SEPARATING NETWORKS
B
C
D
V
TO 4000MC CHANNEL SEPARATING NETWORKS
TO 6000MC CHANNEL SEPARATING NETWORKS
TO 11,000MC CHANNEL SEPARATING NETWORKS
TRANSDUCER CIRCULAR TO RECTANGULAR
SYSTEM SEPARATING NETWORKS
E
F
G
H
TO 4000MC CHANNEL SEPARATING NETWORKS
TO 6000MC CHANNEL SEPARATING NETWORKS
TO 11,000MC CHANNEL SEPARATING NETWORKS

Fig. 3. Schematic diagram of circular waveguide systems for horn-reflector antenna

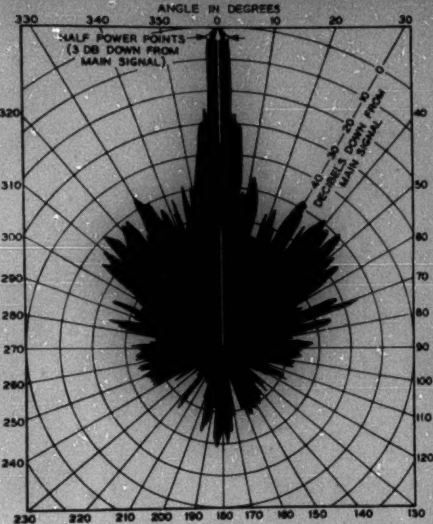


Fig. 4. Horizontal directivity of horn-reflector antenna at 3,740 mc

RETURN LOSS OR VSWR

One of the advantages of this type of antenna is its excellent return loss or low-voltage standing-wave ratio (VSWR) over the very wide frequency band from 3,700 to 11,700 mc. A good impedance match throughout the antenna-waveguide system is important in long-haul radio systems as the waveguide echoes or reflections from an antenna at the end of a long waveguide run may affect the message-handling capacity of the system.

The horn-reflector antenna tapers down with no internal projections to an aperture $11^1/_2$ inches square, to which is attached a precision-cast aluminum feed horn. The internal contour of the 21-inch-long feed horn continued the taper in a gradual hyperbolic transformation designed to match the antenna into a 2.812-inch-ID circular waveguide over the frequency range 3,700 to 11,700 mc. The glass-fiber weather cover across the antenna aperture is sloped to minimize reflections back to the focal point or feed horn. The weather cover is designed to take heavy wind loads without any auxiliary braces or supporting structure, thus minimizing reflections from such sources.

Table IV. Antenna Cross-Polarization Characteristics

Frequency, Mc	XPD, Db
3,740	46
6,325	51
10,960	53

The waveguide is drawn to precision tolerance, and each section is carefully fabricated and tested to ensure that its return loss is 50 db or more (VSWR maximum = 1.0065).

Table V shows the minimum return loss or maximum VSWR measured for a representative antenna of this type with a 2.812-inch-ID circular-waveguide feed across the 4,000-, 6,000-, and 11,000-mc common-carrier bands.

Fabrication

The horn-reflector antenna is shown in more detail in Fig. 6. The antenna is mounted on gimbal supports on a square I-beam mounting frame which in turn swivels on a circular mounting base. Provision is made for ±5-degree adjustments in azimuth and ±3-degree adjustments in elevation angle.

The antenna weighs about 1,700 pounds and is $20^1/_2$ feet high, 11 feet wide, and 9 feet deep. Its focal length is $7^1/_2$ feet, and the front and side angles of the horn are 30 degrees.

The sides of the horn are fabricated of 1/16-inch aluminum alloy reinforced by extruded aluminum "hat section" stiffeners.

The paraboloidal surface of the reflector is held to an accuracy of ±1/16 inch, which, at the highest operating frequency of 11,700 mc, is approximately ±1/16 of a wavelength. The reflector is fabricated in two parts with a dividing line from front to rear. A reflector skin for each half is formed by stretching a 1/8-inch aluminum alloy sheet over a glass-fiber reinforced plastic die having the desired paraboloidal contour. The reflector skins are attached to thick aluminum spars which have been routed to the desired contour. Formed horizontal ribs are attached in the reflector assembly fix-

Table V. Antenna Impedance Characteristics

Frequency, Mc	Polarization	Return Loss, Db	VSWR
3,700 to 4,200	Vertical	43	1.015
	Horizontal	51	1.0055
5,925 to 6,425	Vertical	40	1.020
	Horizontal	48	1.008
10,700 to 11,700	Vertical	49	1.007
	Horizontal	47	1.01

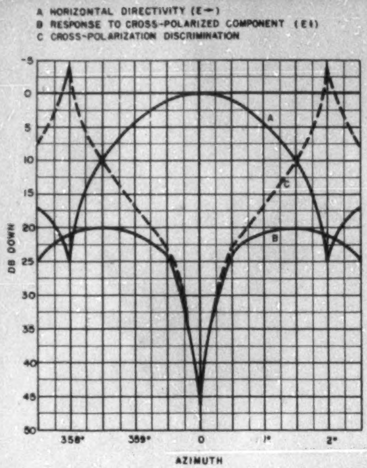


Fig. 5 (left). XPD of main lobe of horn - reflector antenna with azimuth ±2.5 degrees, and 3,740 mc

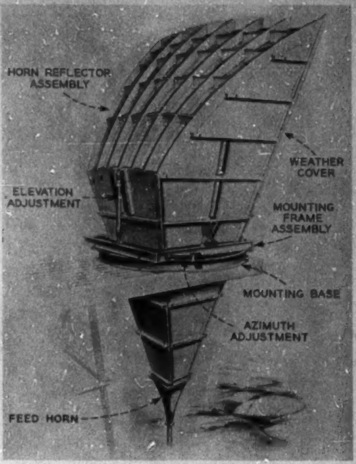


Fig. 6 (right). General view of horn-reflector antenna

ture. Rather extensive tooling was developed to ensure that the reflector surface was kept within tolerance and that the same parts of different antennas were interchangeable in production.

The antenna was designed to withstand heavy snow, ice, and wind loads of 100 pounds per square foot on flat surfaces, equivalent to average wind velocities of about 150 miles per hour. Static load tests up to 100 pounds per square foot were made on a preproduction model of the antenna, and the extensive instrumentation indicated that at this static load no part of the antenna was overstressed. The 64.5-square-foot aperture of the antenna is protected from the weather by a 4-ply polyester impregnated glass-fiber cover 0.040 inch thick.

All seams of the antenna are sealed against weather and radio-frequency leakage by a combination gasket employing nonvulcanized conducting rubber and a fine-woven wire mesh. Though primarily a weather seal, the conducting rubber also provides an important secondary protection against the leakage of microwave energy. The antenna is substantially airtight with a leakage rate of less than 15 cubic feet per hour. The antennas and waveguide at a typical installation are charged with dry air at a pressure of 30 pounds per square foot (0.2 pound per square inch) from a silica gel dehydrator to prevent condensation from affecting the transmission characteristics of the antenna-waveguide system.

The antenna is designed to be shipped from the factory completely assembled and pressure-tested, or packed in four boxes for field assembly at the repeater station site.

The waveguide is a low-phosphorus, high-conducitivity copper waveguide of 2.812±0.003 inches ID and a 3±0.006-inch outside diameter. The ovality or out-of-roundness is held to a difference of 0.004 inch between major and minor diameter at any point along the inside surface of the waveguide. A precision air gage has been developed which makes it possible to insert a measuring head in the waveguide under test and to read directly on an external meter the ID of the waveguide. The measuring device is accurate to ±0.0002 inch throughout its range of 2.812±0.005 inches. Since the XPD is dependent upon the out-of-roundness of the waveguide, the air gage is used by the drawing mill to check each length of tubing drawn. A corrosion-resisting steel flange is "cold-headed" on the copper waveguide, which has had a slight shoulder machined on each end. An O-ring gasket is used between sections of waveguide to ensure a pressure- and weather-tight joint.

The use of the horn-reflector antenna, circular waveguide, and a new design of system and channel-separating waveguide networks promises considerable economies in the expansion of Bell System long-haul radio relay facilities.

References

1. Microwave Repeater Research, H. T. Friis. *Bell System Technical Journal*, New York, N. Y., Apr. 1948.

2. Horn Antenna System, A. C. Beck, H. T. Friis. *U. S. Patent no. 2,416,675*, Mar. 4, 1947.

3. The Ultra-Bandwidth Finline Coupler, S. D. Robertson. *Proceedings*, Institute of Radio Engineers, New York, N. Y., June 1955.

4. Coupled-Wave Theory and Waveguide Applications, S. E. Miller. *Bell System Technical Journal*, May 1954.

5. An Unattended Broad-band Microwave Repeater for the TD-2 Radio Relay System, R. W. Friis, K. D. Smith. *Electrical Engineering* Nov. 1951, pp. 976–81.

Vol.
Pt. 1
1958

Data Transmission Testing Set

J. E. BOUGHTWOOD
ASSOCIATE MEMBER AIEE

T. A. CHRISTIE
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a transmission testing set for high-speed data circuits. The set is an electronic, transistorized device intended primarily for laboratory use in the development of binary data transmission systems.

UNTIL RECENT YEARS the transmission of data in binary form was confined almost exclusively to telegraphic or similar low-speed communication systems. Transmission speed was limited by the rate of operation of printing telegraph or other terminal equipment. Distortion-measuring devices for determining the efficiency of synchronous or start-stop telegraph transmission circuits or equipment were well known to the art.[1,2] The recent introduction of integrated data-processing systems and computer centers requires the transmission of voluminous information in binary form and at high speeds. To develop intelligently circuits and equipment to handle these data required first the development of a transmission measuring tool capable of operation at the high speeds involved. It is the purpose of this paper to describe an all-electronic, transistorized testing set which will measure bias distortion, characteristic distortion, and peak fortuitous distortion at transmission speeds from 200 to 1,400 bits per second.

Basic Principles

Distortion in binary pulse transmission may be defined as the deviation in length or relative time relationship of the received pulses from the length and time relationship of their transmission. A signal is defined as unbiased when the unit-marking and spacing pulses are of the same length or time duration. Bias distortion usually results from improper adjustment of terminal or circuit equipment. A means for measuring this type of distortion should be included in a data transmission testing set.

Distortion caused by electrical or magnetic characteristics of the equipment or circuit under test is defined as characteristic distortion. It is the amount by which a given pulse is lengthened, shortened, or displaced as a result of the presence or absence of neighboring pulses, and is caused by "memory" effects in the transmission path. More exactly, the maximum amount by which the space-to-mark and mark-to-space transitions of a selected pulse are shifted from their ideal locations by the presence or absence of neighboring pulses is a measure of the characteristic distortion. A device to measure characteristic distortion might then consist of a pulse pattern generator for sending various combinations of marking and spacing pulses, a circuit for selecting one of the received pulses for examination, and a circuit for comparing the distorted received pulse with the undistorted transmitted pulse to indicate the amount of distortion. The characteristic distortion would be taken as the largest value indicated for all possible combinations of pulses from the pattern generator.

Fortuitous distortion is that displacement of signal transitions resulting from extraneous electrical interference in the transmission circuit. This interference is statistical in nature and has a random relation to the testing signals. The "instantaneous" distortion caused by fortuitous interference will vary widely although the "average" effect on a repeated pulse will integrate to zero. It is conventional to measure fortuitous distortion as the largest transition displacement encountered in a given period. The measurement of fortuitous distortion should not be influenced by the characteristic distortion of the circuit under test. A device to measure fortuitous distortion might then consist of a pattern generator for sending alternate marking and spacing pulses of equal length, a circuit for selecting one of the pulses for examination, and a peak-reading instrument for measuring the largest transition displacement during the period of observation.

Description

In contrast to telegraph codes, data codes have not yet been standardized as to transmission speeds or number of code elements per character. For the design of a testing set, a code of ten elements and speeds from 200 to 1,400 bits per second was selected arbitrarily.

A block diagram of the testing set is shown in Fig. 1. It is completely transistorized except for two magnetron beam-switching tubes and a subminiature electrometer tube. Both transmitter and receiver sections are controlled by the output of a 200- to 1,400-cycle sine wave oscillator adjustable in steps of 200 cycles. The oscillator frequency enters the transmitter section through a squaring amplifier. The square wave from this amplifier is then differentiated and the resultant pulses used to trigger a standard bistable flip-flop stage. Because the flip-flop reverses once per cycle of the applied frequency, its output is a highly symmetrical or unbiased square wave, i.e., the time duration of the negative and positive loops are equal. Two appropriately phased outputs from the flip-flop drive the odd and even grids respectively, of a 10-position beam-switching tube to advance the beam one position for each reversal of the flip-flop. A 10-position signal pattern selector switch associated with the beam-switching tube permits any one of ten signal patterns to be selected for transmission at a bit rate equal to the oscillator frequency in cycles per second.

The adjustable oscillator also operates a second beam-switching tube in the receiver portion of the test set so that both tubes run at exactly the same speed. The drive circuit for the receiver tube includes a continuously adjustable phase shift network of a 360-degree range. This network, in conjunction with a phase switch associated with the beam-switching tube, permits a gate pulse and a reference pulse, to be obtained once per revolution, which may be placed in any desired time relationship with the transmitter signal pattern. These pulses have precisely the same length as the transmitter pulses. The gate pulse selects the leading or trailing portion of a given signal pulse to be measured by means of the coincidence circuit. The output of the coincidence circuit is applied to the characteristic distortion meter circuit, where it is compared with the reference pulse for characteristic or bias distortion measurements. To measure peak fortuitous distortion, the output of the coincidence circuit is applied to the fortuitous distortion meter circuit.

Ten signal patterns each consisting of ten bits are available for transmission, as indicated in Fig. 2. The polarity of transmission may be shown or reversed

---

Paper 58-172, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing November 19, 1957.

J. E. Boughtwood and T. A. Christie are with The Western Union Telegraph Company, New York, N. Y.

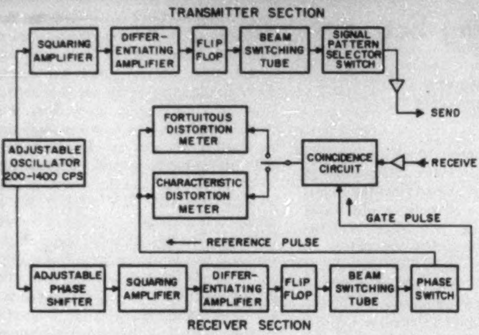
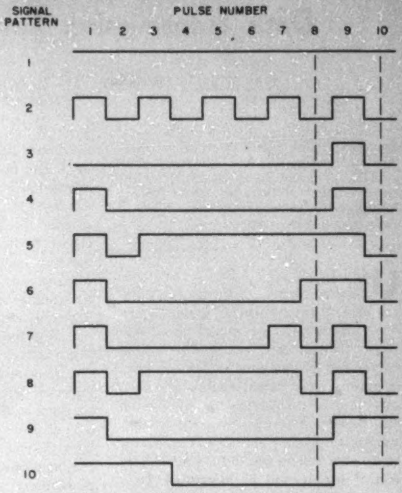

TRANSMITTER SECTION
SQUARING AMPLIFIER
DIFFER-ENTIATING AMPLIFIER
FLIP FLOP
BEAM SWITCHING TUBE
SIGNAL PATTERN SELECTOR SWITCH
SEND
FORTUITOUS DISTORTION METER
ADJUSTABLE OSCILLATOR 200-1400 CPS
COINCIDENCE CIRCUIT
RECEIVE
CHARACTERISTIC DISTORTION METER
GATE PULSE
REFERENCE PULSE
ADJUSTABLE PHASE SHIFTER
SQUARING AMPLIFIER
DIFFER-ENTIATING AMPLIFIER
FLIP FLOP
BEAM SWITCHING TUBE
PHASE SWITCH
RECEIVER SECTION

Fig. 1. Block diagram of data transmission testing set


SIGNAL PATTERN
PULSE NUMBER
1
2
3
4
5
6
7
8
9
10

POLARITY OF MARKING AND SPACING REVERSIBLE

Fig. 2 (right). Signal patterns available from transmitter

as desired. After transmission through the circuit or equipment under test, pulse *9* is normally used for characteristic distortion analysis. The signal patterns have been selected to include combinations producing the maximum distortion of pulse *9* that could occur with normal random message traffic.

Bias and Characteristic Distortion Measurement

A circuit for the selection of the leading or trailing portions of a received signal pulse for bias or characteristic distortion measurement is shown in Fig. 3. After transmission over the system under test, the signals enter a coincidence circuit at *A*. A gate pulse from the receiver beam-switching tube is selected by means of the phase switch, and it enters at *B*. Transistors *Q1* and *Q2* comprise a coincidence circuit, and transistors *Q3*, *Q4*, and 50-0-50 scale meter comprise a differential metering circuit. The output of the coincidence circuit enters the metering circuit through transistor *Q3* and a reference pulse from the receiver beam-switching tube enters through transistor *Q4*. Because the pulses are applied differentially to the meter, a zero reading will result when the pulses have equal areas. The manner in which the amount of bias or characteristic distortion is indicated is illustrated in Fig. 4. Bias measurements are made on a-c signals, signal pattern *2*, position *2*, of the signal pattern selector switch. Any pulse of the pattern may be selected for bias distortion measurements but pulse *9* is used for characteristic distortion measurements. It is necessary to measure the displacement of the leading transition and the trailing transition of the pulse with respect to their undistorted positions, in order to determine the amount of distortion. These transitions are measured separately, as indicated in Fig. 4. The received a-c signal pattern is shown at *A*. If these signals are unbiased they will have exactly the same length as the gate pulse from the receiver beam-switching tube shown at *B*. The gate pulse may be made to straddle either the leading or the trailing edge of a received pulse by means of the phase switch and the adjustable phase shifter included in the receiver beam-switching tube drive circuit.

The output of the coincidence circuit is shown at *C* and the reference pulse at *D*. The length of the *D* pulse is exactly the same as that of the *B* pulse but its amplitude is made to be precisely one half the amplitude of the *C* pulse. By properly adjusting the phase of gate pulse *B*, the *C* pulse for the leading transition of signal pulse *9* can be made exactly the same length as the *C* pulse for the trailing transition of the pulse *9* so that both *C*

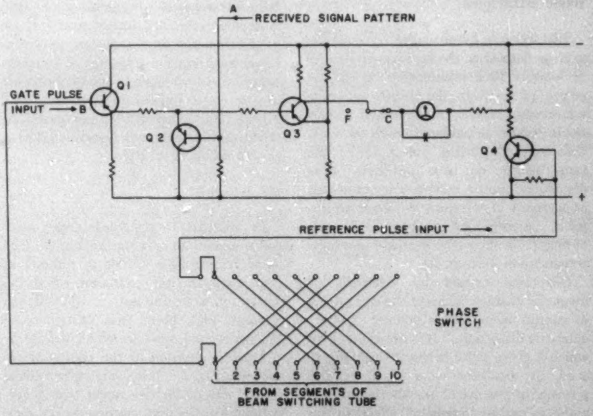

A
RECEIVED SIGNAL PATTERN
GATE PULSE INPUT → B
Q1
Q2
Q3
F
C
Q4
−
+
REFERENCE PULSE INPUT
PHASE SWITCH
1 2 3 4 5 6 7 8 9 10
FROM SEGMENTS OF BEAM SWITCHING TUBE

Fig. 3. Bias and characteristic distortion-measuring circuit

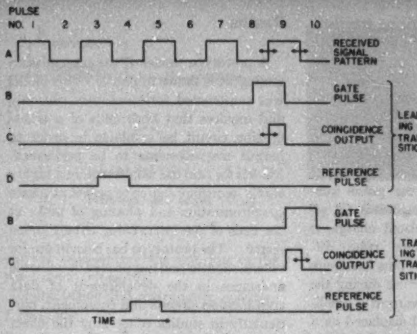


Fig. 4. Theory of bias and characteristic distortion measurements

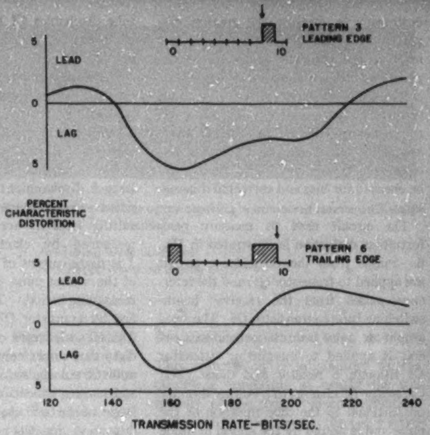


Fig. 5 (right). Typical characteristic distortion curves

pulses have the same length and area. If the signals are unbiased, both *C* pulses are also equal in area to reference pulse *D*. Since the *D* pulse and a *C* pulse are applied alternately and differentially to the zero center d-c meter, a zero reading of the meter will result. If the received signals at *A* are biased, the length of the *C* pulses for the leading and trailing edges will still be equal. However, the *C* pulses will be greater or smaller in area than the *D* pulse, depending on whether the bias is marking or spacing, and the differential meter will indicate directly the direction and magnitude of the bias in per cent of the duration of an undistorted pulse.

An integrating condenser in the differential meter circuit produces a steady meter reading by eliminating the effects of fortuitous interference. A bias measurement is made by alternately switching between two adjacent positions of the phase switch, to select alternately the leading and trailing transitions, simultaneously adjusting the phase shift network for the same meter reading.

Before characteristic distortion measurements can be made, all bias should be removed from the signals to be measured, by the procedure just described to indicate the unbiased condition. As previously stated, characteristic distortion is measured by selecting signal pulse *9*, and observing the displacement of the leading and trailing transitions for the various signal patterns. The displacements are measured relative to the location of the leading and trailing transitions for signal pattern *2* (a-c).

Pulse *9* is selected for measurement by transmitting signal pattern *3*, consisting of a single pulse in position *9*. If gate pulse *B*, Fig. 4, is coincident with any other pulse position than position *9*, the differential meter will read −50% distortion, since only reference pulse *D* is applied to the meter circuit. The leading edge of pulse *9* is located by rotating the phase switch in a clockwise direction until the meter reads other than −50%.

The meter reading is a direct measurement of the leading edge displacement in per cent of the length of an undistorted pulse. This reading may be converted to time if desired since the pulse transmission rate is accurately known. If the leading edge is not affected by characteristic distortion, it will have the same location as for pattern *2* (a-c), and the meter will read zero. The displacement of the trailing edge may similarly be measured by rotating the phase switch to the next adjacent clockwise position. In such fashion the displacement of the transitions are measured for signal patterns *3* through *10*. The amount of characteristic distortion for the particular transmission speed is taken as the largest meter reading of the series. To delineate this distortion characteristic completely, measurements must be made for a range of speeds. Typical curves of distortion versus speed measured in a medium-speed data channel are given in Fig. 5. The effect of speed on the leading transition of pulse *9* for pattern *3*, and the trailing transition of pulse *9* for pattern *6* are clearly shown.

FORTUITOUS DISTORTION MEASUREMENT

Fortuitous distortion displaces the leading and trailing transitions of a signal pulse in a random manner. It is

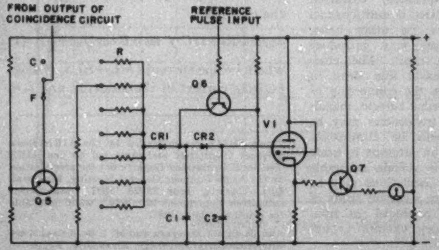


Fig. 6 (left). Peak fortuitous distortion measuring circuit

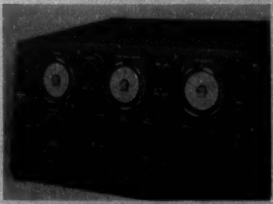


Fig. 7 (right). Data transmission testing set, front view customary to measure the greatest displacement occurring in a specified time interval and to define this value as the peak fortuitous distortion of the circuit or equipment under test. Fortuitous distortion measurements are made on a-c signals, signal pattern 2, to eliminate any meter deflection caused by characteristic distortion. The signals must also be checked for bias and corrected if necessary.

The circuit used to measure peak fortuitous distortion is illustrated in Fig. 6. Pulses from the coincidence circuit are applied to transistor *Q5*, and the reference pulses from the receiver beam-switching tube to transistor *Q6*. The fixed amplitude pulse from the coincidence circuit is applied to integrating capacitor *C1* through a resistor and diode *CR1*. Capacitor *C1* charges to a voltage directly proportional to the time duration of the pulse and is bound by the action of diode *CR1*. Capacitor *C2*, which is small compared to *C1*, charges through diode *CR2* to the same value as *C1*. The reference pulse from the receiver-switching tube discharges *C1* by driving transistor *Q6* to conduction leaving a bound charge on *C2*.

If the next pulse from the coincidence circuit is longer than the preceding pulse because of fortuitous distortion, capacitor *C1* will charge to a higher value, as will capacitor *C2*. In this manner a charge is accumulated on *C2* proportional to the largest displacement of the pulse transition. The voltage of capacitor *C2*, resulting from its accumulated charge, is measured by electrometer tube *V1*. The plate current of *V1* is thus a measure of the widest pulse occurring during the measuring period. This current is amplified by transistor *Q7* and displaced on a 0 to 40 scale meter calibrated to indicate distortion in per cent of the length of an undistorted pulse. Both the leading and trailing transitions are measured, the larger of the two being the peak fortuitous distortion for the particular speed of transmission. Calibration is maintained for different speeds by the automatic selection of the proper charging resistor *R* for a given signaling speed.

Results

A prototype model of the transmission testing set is shown in Fig. 7. This model was constructed for laboratory use only and requires that both ends of a transmission circuit be available in order to permit measurements to be performed. Models for field use would require a highly stable frequency source to permit accurate synchronization and phasing of units at the ends of the intervening circuit being tested. The prototype has been in use for several months and has proved of valuable assistance in the development of data transmission circuits and equipment, particularly in studies concerning the effect of envelope delay distortion on high-speed data transmission.

References

1. TELEGRAPH TRANSMISSION TESTING MACHINE, F. B. Bramhall. *AIEE Transactions*, vol. 50, June 1931, pp. 404–13.

2. A PORTABLE TELEGRAPH BIAS AND DISTORTION MEASURING INSTRUMENT, W. D. Cannon. *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 401–07.

Subharmonics in Iron-Cored Reactor Oscillatory Circuits

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

M. I. EL-KARAKSI
NONMEMBER AIEE

Synopsis: This report gives a careful study of subharmonic sustained oscillations found in iron-cored reactor oscillatory circuits. The analysis is based on the power balance principle, giving satisfactory explanation and understanding of the idea of frequency conversion. The analysis is valid for both the harmonic and subharmonic cases. The ranges of synchronization, the methods of starting, and the conditions of stability of these oscillations are given. Numerical examples for the third subharmonic oscillations are presented. Just outside the locking range, the amplitude and phase of the subharmonic oscillations are no longer constant, but variable; the oscillations are defined as "pulled oscillations." A theoretical analysis for determining the instantaneous phase angle is given. An experimental investigation to test the validity of the theoretical analysis is carried out. The experimental work includes steady and transient conditions, a harmonic source of voltage being applied. The ranges of the supply voltage switching angles required for starting sustained oscillations for different circuit parameters are experimentally obtained. An initial charging of the condenser or an instantaneous application of d-c flux improves the starting conditions.

Subharmonic oscillations generated in oscillatory circuits containing iron-cored reactors have found increasing use in electrical engineering. The production of the 20- or $16^2/_3$-cycle ringing current from the 60- or 50-cycle power respectively, is a commercial application. A static frequency divider which is well stabilized and easily operated is suggested; frequency standardization is a good application of such a frequency divider. When operated at high supply frequencies, sustained oscillations may be utilized in multichannel carrier telephony. On the other hand, subharmonic oscillations may introduce difficulties in certain cases. The compensation of transmission line drop in power circuits requires the connection of series capacitors, abnormal current magnitudes of submultiple frequencies may be produced; these should be eliminated. In general, whenever an attempt is made to tune circuits which include saturable reactors, subharmonic oscillations may be encountered. Therefore a careful study of sustained oscillations produced in iron-cored reactor oscillatory circuits is important.

THE circuit to be analyzed is the simple series oscillatory circuit shown in Fig. 1, where ϕ represents the flux linkage of the iron-cored reactor, including any leakage flux; R is a load resistance including the winding resistance of the reactor coil; and C is a capacitor. A harmonic source of voltage $e_s[=E\cos(\omega_s t+\alpha)=E\cos(m\omega t+\alpha)]$ is applied to the circuit, where $\omega_s/2\pi=m\omega/2\pi=$ supply frequency, and m is an integer. Both hysteresis and eddy current losses may be approximately represented by a constant resistance r shunting the inductor.

Fig. 2 shows the d-c $\phi\rightarrow i_\phi$ characteristic of the iron-cored reactor used. It was found more appropriate to represent the characteristic by a power series; thus

$$i_\phi=a_1\phi+a_3\phi^3+a_5\phi^5+\ldots$$

The differential equation for the current in the circuit is

$$d\phi/dt+Ri+1/C\int i\,dt=E\cos(\omega_s t+\alpha) \quad (1)$$

where $i=$ line current $(=i_\phi+i_r)$, $i_\phi=$ magnetizing current of the inductor, and $i_r=$

Paper 57-745, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted September 13, 1954; made available for printing July 9, 1957.

ABD EL-SAMIE MOSTAFA and M. I. EL-KARAKSI are with Alexandria University, Alexandria, Egypt.

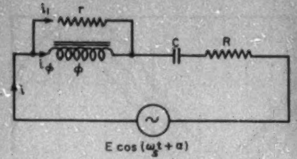


Fig. 1. Ferro-resonant series circuit

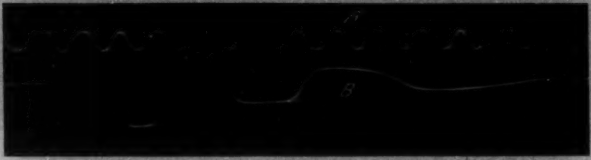


Fig. 3. Oscillogram giving free decayed charge across capacitor in iron-cored reactor oscillatory circuits

A—Timing wave 50 cycles
B—Charge waveform during decay

current representing the iron loss in the reactor $= (1/r)(d\phi/dt)$.

Substituting and rearranging, equation 1 becomes

$$d\phi/dt + E_{eq} i_\phi + (1/C_{eq}) \int i_\phi dt = E_{eq} \cos(\omega_s t + \alpha) \quad (2)$$

where

$$R_{eq} = [1/(1+R/r)][R + (\phi/i_\phi)/rC]$$
$$C_{eq} = C[1+R/r]$$
$$E_{eq} = E/(1+R/r)$$

To produce sustained subharmonic or harmonic oscillations, energy must be transferred from the voltage source to the oscillation at the right instant and in the right direction; in other words, the oscillation must be in some form of synchronization with the supply. If the supply voltage e_s is absent, damped free oscillation takes place with a frequency which decreases as the amplitude decreases, the condenser being assumed initially charged; see Figs. 3, 4, and 5. However, if e_s is applied, both free and forced oscillations take place. If the frequency of the free oscillation, being dependent on the amplitude, is near $n\omega/2\pi$ (n being an integer), energy is transferred from the voltage source to the free oscillation resulting in disturbing the phase angle between the oscillation and the applied voltage. If a state is reached in which the phase balance of the oscillation can be maintained at the frequency $n\omega/2\pi$, the oscillation can be sustained. Such sustained oscillation is therefore synchronized oscillation.

Neglecting the harmonics of both forced and synchronized oscillations, the flux ϕ of equation 2, at the steady state, may be assumed to be of the following form

$$\phi = \phi_m \cos m\omega t + \phi_n \cos(n\omega t + \theta) \quad (3)$$

This assumption is satisfactory as the circuit is resonating at nearly the frequency of the synchronized oscillation. Substituting equation 3 in the $i_\phi \rightarrow \phi$ series gives

$$i_\phi = \Sigma K_{k,\pm h} \cos[(km \pm hn)\omega t \pm h\theta] \quad (4)$$

In equation 4, the summation is made for $k=0, 1, 2, 3$, etc., $h=0, 1, 2, 3$, etc., k different from h, and $(k \pm h)$ odd. The upper limit of the summation depends on the number of terms in the $i_\phi \rightarrow \phi$ series. The constants $K_{k,\pm h}$ are functions of ϕ_m, ϕ_n as well as coefficients a_1, a_3, a_5, etc.

Transfer of Active Power Through the Coil

As the power in the coil is equal to the mean value of $\int v i_\phi \, dt$, and v is the voltage across the reactor and equal to $d\phi/dt$, the current components responsible for the active power P_c in the coil must be of the form $[M \sin m\omega t + N \sin(n\omega t + \theta)]$, which can be selected from equation 4. Carrying out this process gives

$$P_c = P_{c_m} + P_{c_n} \quad (5)$$

where

P_{c_m} = power given to coil at the supply frequency $m\omega/2\pi$

$$= -\frac{1}{2} m\omega\phi_m \sum_{(k-1)m = \mp hn} \mp K_{k,h} \sin h\theta \quad (6)$$

and

P_{c_n} = power given to coil at the synchronized oscillation frequency $n\omega/2\pi$

$$= -\frac{1}{2} n\omega\phi_n \sum_{km = (1 \mp h)n} \mp K_{k,h} \sin(h \mp 1)\theta \quad (7)$$

With equations 6 and 7, it may be shown that the sum of P_{c_m} and P_{c_n} is zero. This conclusion is natural since the coil dissipates no energy. The coil derives

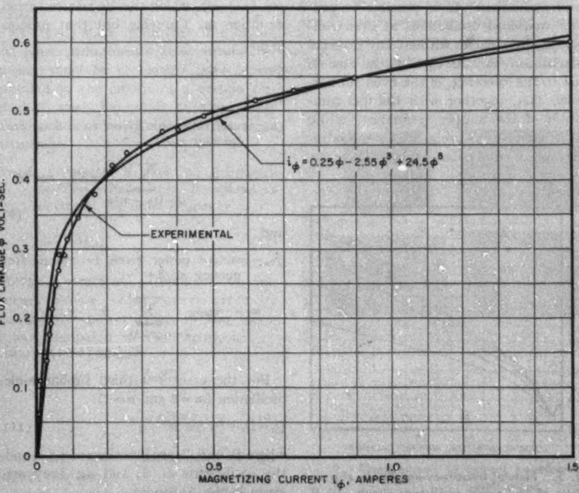


Fig. 2. Representation of magnetization curve

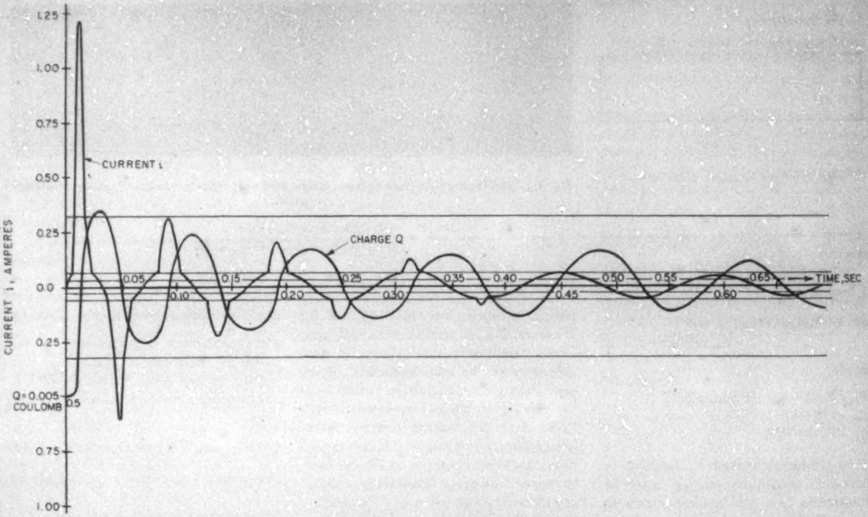


Fig. 4. Theoretical current and charge waveforms during decay in iron-cored reactor oscillatory circuit. C=100 μf, R=10 ohms, initial charge on condenser=0.005 coulomb

power from the supply at its frequency $m\omega/2\pi$ and delivers the same amount of power at another frequency $n\omega/2\pi$, i.e., the coil acts as a frequency converter.

If only three terms in the $i_\phi \rightarrow \phi$ series are considered, the expressions for P_{c_m} and P_{c_n} for different values of m and n may be easily obtained from equations 6 and 7 and are given by

1. Third subharmonic oscillation, $m=3$ and $n=1$:

$$P_{c_3}=D_1 \sin 3\theta=-P_c \tag{8}$$

2. Fifth subharmonic oscillation $m=5$ and $n=1$:

$$P_{c_5}=D_2 \sin 5\theta=-P_{c_1}$$

3. Third harmonic oscillation $m=1$ and $n=3$:

$$P_{c_1}=D_3 \sin \theta=-P_{c_3}$$

4. Second subharmonic oscillation $m=2$ and $n=1$:

$$P_{c_2}=D_4 \sin 4\theta=-P_{c_1}$$

where D_1, D_2, D_3, and D_4 are functions of ϕ_1, ϕ_3, and the coefficients a_1, a_3, and a_5 given in the Appendix. It is clear that maximum power converted in coil for these four cases takes place when $3\theta=\pi/2$, $5\theta=\pi/2$, $\theta=\pi/2$, and $4\theta=\pi/2$ respectively.

In general, the converted power depends on the phase angle θ between the flux having the supply frequency and that of the sustained oscillation. This phase angle depends on the amount of detuning between the natural frequency of the circuit and the frequency of the synchronized oscillation. It is possible to generate sustained oscillations at even submultiple frequencies without any external polarization to the core as in case 4. Due to the presence of the even subharmonic flux, together with the flux component of the supply frequency, a d-c flux component is produced, which is responsible for the polarization of the core.

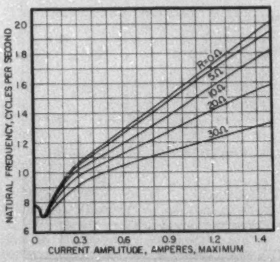


Fig. 5. Natural frequency-amplitude characteristic of iron-cored oscillatory circuit with R as parameter

Reactive Power in Coil

It is clear that the current components responsible for the reactive power Q_c in the coil must be of the form $A \cos m\omega t+B \cos (n\omega t+\theta)$, which can be selected from equation 4. Carrying out that process gives

$$Q_c=Q_{c_m}+Q_{c_n}$$

where

Q_{c_m} = reactive power given to coil at frequency $m\omega/2\pi$

$$=-\frac{1}{2} m\omega\phi_m \sum_{(h-1)m=\mp hn} \pm K_{k,h} \cos h\theta \tag{9}$$

and

Q_{c_n} = reactive power given to coil at frequency $n\omega/2\pi$

$$=-\frac{1}{2} n\omega\phi_n \sum_{km=(1\mp h)n} \pm K_{k,h} \cos (h\mp 1)\theta \tag{10}$$

For the case of a third subharmonic oscillation ($m=3$ and $n=1$)

$$Q_c=D_5+D_6 \cos 3\theta \tag{11}$$

where D_5 and D_6 are functions of ϕ_1, ϕ_3 and the coefficients a_1, a_3 and a_5; they are given in the Appendix.

As stated before, power converted in

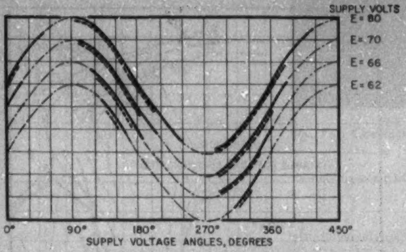


Fig. 6. Ranges of supply voltage switching angles at which third subharmonic sustained oscillations may be started for different values of resistance R and supply voltage E, C=100 μf, —— R=10 ohms, - - - R=20 ohms

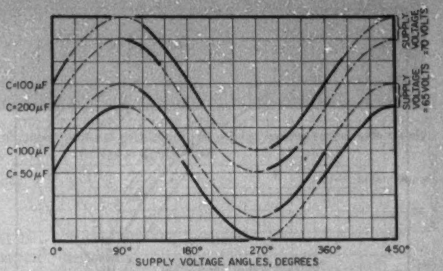


Fig. 7. Ranges of supply voltage switching angles at which third subharmonic sustained oscillations may be started for different values of capacitance C and supply voltage E. R=10 ohms

coil is maximum when $3\theta=\pi/2$, in which case the amount of detuning is equal to zero. If there is some detuning in the system, there must be an additive reactive power taken from the supply through the conversion in the coil in order to keep the system in a synchronized state. This reactive power constitutes the terms containing $\cos 3\theta$ in equation 11.

Quantitative Evaluation of Sustained Oscillation

With the conception that the coil could be considered as a frequency-converter-source, quantitative evaluation of the sustained oscillations may be determined.

At steady-state conditions, the different power components at a certain frequency must be in balance, i.e., power supplied to the circuit at a certain frequency must be equal to the power dissipated in the circuit at the same frequency.

Power generated in coil at frequency $n\omega/2\pi=-P_{c_n}=$power dissipated in circuit at the same frequency

$$=\frac{1}{2}I_n^2R_{eq} \tag{12}$$

where $I_n=$current component of frequency $n\omega/2\pi$. Also, power given by the supply=power given to the coil+power dissipated in the circuit, i.e.,

$$\frac{1}{2}E_{eq}I_m\cos\alpha=P_{c_m}+\frac{1}{2}I_m^2R_{eq}$$

$$=\frac{1}{2}(I_m^2+I_n^2)R_{eq} \tag{13}$$

where $I_m=$current component of the supply frequency.

Again, for steady-state operation, the sum of the reactive powers in the circuit at a certain frequency must vanish; therefore reactive power at frequency $n\omega/2\pi=$reactive power in coil+reactive power in condenser

$$=Q_{c_n}-\frac{\frac{1}{2}I_n^2}{n\omega C_{eq}}=0 \tag{14}$$

and reactive power at a frequency $m\omega/2\pi=0$, giving reactive power in coil+reactive power in condenser=reactive power delivered by the supply, i.e.

$$Q_{c_m}-\frac{\frac{1}{2}I_m^2}{m\omega C_{eq}}=\frac{1}{2}E_{eq}I_m\sin\alpha \tag{15}$$

Using equations 12, 13, 14, and 15, together with the relations of R_{eq}, C_{eq}, and E_{eq}, the values of ϕ_m, ϕ_n, θ, and α may be obtained for given values of E, C, R, and a certain iron-cored reactor.

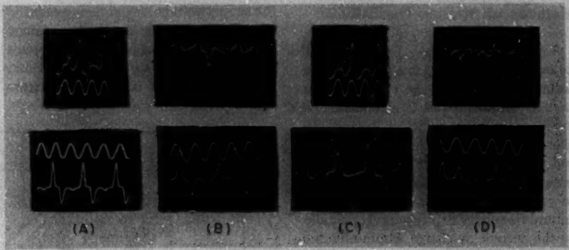
Conditions of Stability of Oscillations

With the aid of equations 12, 13, 14, and 15, the stationary amplitudes and phase of the sustained oscillation may be obtained; whether these values occur in practice or not depends on their stability.

Active power at a frequency $(n\omega/2\pi)$ is derived from the coil through convetsion, as stated before, and is equal to $-P_{c_n}$. Also, power lost in circuit at the same frequency $n\omega/2\pi$ is equal to $(1/2)I_n^2R_{eq}$. At the stationary state, these two powers are equal; see equation 12. However, due to any slight disturbance in the system, the balance in power is disturbed and there will be a net power. Thus

$$-P_{c_n}-(1/2)I_n^2R_{eq}=\text{net power at a frequency } n\omega/2\pi=P_{net} \tag{16}$$

For the oscillation to be stable, any



	A	B	C	D
ORDER OF SUBHARMONIC	SECOND	THIRD	FOURTH	SIXTH
SUPPLY VOLTAGE, VOLTS, RMS	60	69	53	100
CIRCUIT RESISTANCE, OHMS	15	6	5	5.2
CIRCUIT CAPACITANCE μF	60	60	100	200
CIRCUIT CURRENT, AMPERES	0.84	1.3	1	1.6

Fig. 8. Oscillograms of current waveforms of second, third, fourth, and sixth subharmonic oscillations. Supply voltages indicated

Top row: direct current, amperes. A—1, B—0.5, C—1, D—0.8. Bottom row: no direct current

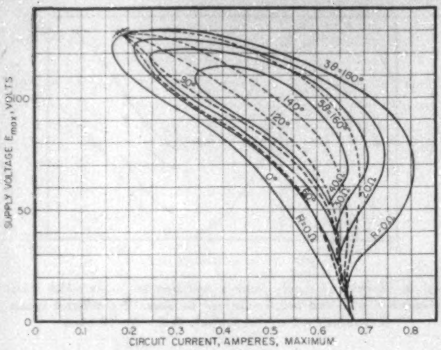


Fig. 9. Theoretical synchronizing ranges of third subharmonic sustained oscillations

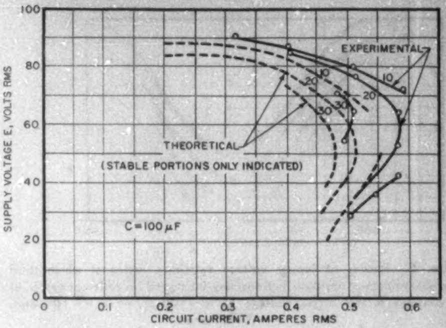


Fig. 10. Experimental and theoretical synchronizing ranges of third subharmonic oscillation

slight increase ΔI_n, in its stationary amplitude, must be accompanied by a decrease ΔP_{net} in the net power; otherwise the amplitude would rise indefinitely. Therefore the condition of stability is that $\Delta P_{net}/\Delta I_n$ must be negative at the stationary values, i.e.,

(dP_{net}/dI_n) (at the stationary values)<0, or
$(d(P_{net}/I_n^2)/dI_n)$ (at the stationary values) <0

Using equation 16 gives

$$d(P_{net}/I_n^2)/dI_n = d(-P_{c_n}/I_n^2)/dI_n$$

Substitution in the inequality gives the condition of stability, namely

$$[d(-P_{c_n}/I_n^2)dI_n] \text{ (at the stationary values)}<0 \qquad (17)$$

Methods of Starting Oscillations

A free oscillation may be sustained if its amplitude attains such a value that the natural frequency of the circuit is near $n\omega/2\pi$. Although this is a fundamental condition, it is not sufficient because the phase angle between the free oscillation and the supply voltage must be within certain limits in order that the former may be pulled in and hence sustained. The natural frequency depends fundamentally on the current or charge amplitude; see Figs. 3, 4, and 5. Therefore the initial amplitude of oscillation is an important factor in determining the possibility of producing sustained oscillations. This amplitude depends on initial conditions, namely

1. Supply voltage switching angle, i.e., the magnitude of the supply voltage at the instant of energizing the circuit.
2. Initial charge on the condenser.
3. Initial flux linkage of the iron-cored inductor.

There is a range of supply voltage switching angles in which the free oscillations may be synchronized and hence sustained; see Figs. 6 and 7. These last two figures are practically obtained by closing the circuit at certain supply voltage switching angles with the use of a synchronous-relayed switch; the relay being preset so that the switch permanently closes the circuit when the current reaches the value of the required sustained oscillation. It is clear from Fig. 6 that, for a given capacitance, the range of the supply voltage switching angles increases as the supply voltage increases and as the circuit resistance decreases. Therefore, for starting by direct switching, the supply voltage should be increased to the maximum value at which sustained oscillations are possible and the circuit resistance decreased to the minimum possible value. It is again clear from Fig. 7 that for the same supply voltage the range of the supply voltage switching angles may be increased by reducing the capacitance. These characteristics of the circuit may be analyzed as follows.

For the same supply voltage and capacitance, an increase of the resistance leads to a decrease of the first maximum amplitude of oscillation, and hence the range of supply voltage switching angles decreases. Increasing the supply voltage is in certain respects equivalent to a decrease in circuit resistance, and hence the increase of supply voltage will increase the supply voltage switching angles. For the same supply voltage and the same resistance, an increase of capacitance necessitates a reduction in the effective inductance of the circuit to have a third subharmonic oscillation; the first maximum amplitude of oscillation must therefore be increased, and hence the range of the supply voltage switching angles decreases.

The range of the supply voltage switch-

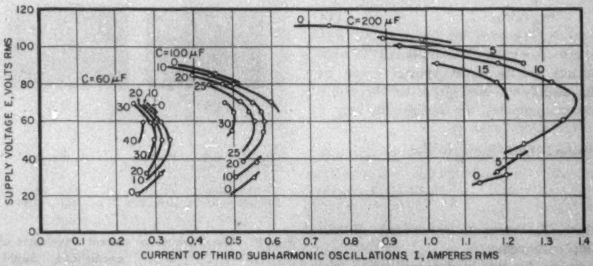


Fig. 11. Experimental voltage-current synchronizing ranges of third subharmonic oscillations

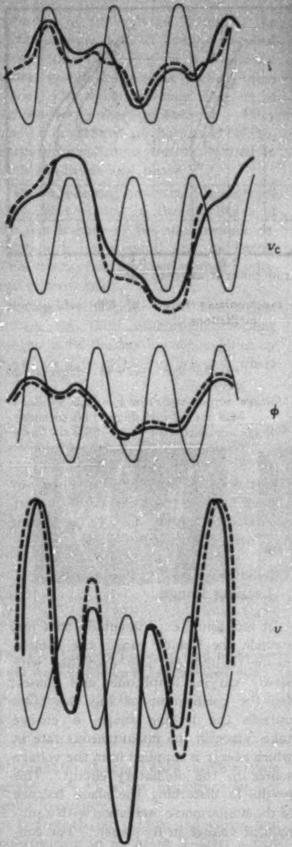


Fig. 12. Theoretical waveforms together with the experimental oscillograms of i, v_c, ϕ, and v. Theoretical waveforms are shown dashed. Sinusoidal supply voltage also indicated

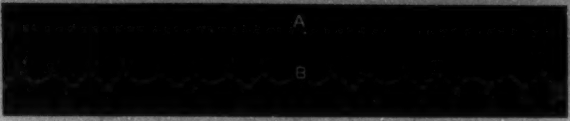


Fig. 13. Oscillogram of current waves of oscillations just outside locking range, with supply voltage as timing wave

A—Timing wave 50 cycles

B—Current wave

ing angles may be increased by either charging the condenser or applying instantaneously a polarizing flux prior to energizing the circuit. In this way the first maximum amplitude of the charge can attain any value independent of the supply voltage switching angle. Experiments showed that, with a reasonable initial charge, 1 out of 50 switching operations fails to produce sustained oscillations at one third of the supply frequency. Starting by applying instantaneously a d-c flux yields even better results. If the d-c flux is not only instantaneously applied but permanently kept in the core, sustained oscillations, especially those of even subharmonics (one half and one fourth the supply frequency, for example) are produced whatever the switching angle may be. With permanent polarizing flux linkage, the frequency of the sustained oscillations may be varied with variation of the magnitude of the polarizing current. Sustained oscillations of frequencies one fifth, one fourth, one third, and one half the supply frequency are obtained experimentally by only varying the polarizing current. In this way, the system is considered as a variable frequency divider. As a practical application, a static, easily operated and stabilized frequency divider is suggested operating at different spot frequencies.

It should be noted here that the waveforms of the sustained oscillation of a certain order (one third the supply frequency, for example) are independent of the supply voltage switching angle, the circuit parameters (R, C, and E) being the same; they are generally independent of the method of starting whether it is initially charging the condenser or applying instantaneously a d-c flux in the core. However, in case there is a permanently polarizing d-c flux the waveforms will not be skew symmetrical. Fig. 8 shows several oscillograms for the second, third, fourth and sixth subharmonic currents with and without a polarizing flux. The first maximum amplitude of the charge on the condenser depends on the supply voltage switching angle, the condenser being initially uncharged.

Numerical Calculations for Sustained Oscillations of One Third the Supply Frequency

One of the cases investigated experimentally will be calculated. The ranges of synchronization, the limiting values of the supply voltages, and the conditions of stability will be estimated. The magnetization curve of the iron-cored reactor used is that shown in Fig. 2, and which is represented by the following equation

$$i_\phi = 0.25\phi - 2.55\phi^3 + 24.5\phi^5 \qquad (18)$$

where i_ϕ is the current in amperes, ϕ the flux linkage in volt-seconds, and the representation is valid up to $i_\phi = 1.5$ amperes.

It is found experimentally that the iron loss resistance r has an average value of 2,750 ohms while the possible ranges of the series resistance R are up to 60 ohms, and the quantity $(1+R/r)$ is very near unity; therefore E and C can stand satisfactorily for E_{eq} and C_{eq} respectively. To simplify calculations, the average value of ϕ/i_ϕ, in the whole operating ranges, is used instead of ϕ/i_ϕ in the expression of R_{eq}. Again, it was found that this average value is approximately equal to 1.0 and thus $(\phi/i_\phi)/rC$ is small and R can stand satisfactorily for R_{eq}.

Assume the initial conditions and the different circuit parameters which are C, R, and E to be such that the sustained oscillations are of one third the supply frequency. Let the flux linkage be of the form

$$\phi = \phi_3 \cos 3\omega t + \phi_1 \cos(\omega t + \theta)$$

where $\omega/2\pi$ = frequency of the sustained oscillation, and $3\omega/2\pi$ = supply frequency.

Substitution in equation 18 gives

$$i_\phi = A_1 \cos 3\omega t + A_2 \sin 3\omega t + B_1 \cos(\omega t + \theta) + B_2 \sin(\omega t + \theta) + \text{components of frequencies other than } 3\omega/2\pi \text{ and } \omega/2\pi \qquad (19)$$

where A_1, A_2, B_1, and B_2 are functions of ϕ_1, ϕ_3, and θ, and given in the Appendix.

From equation 8

$$P_{e_3} = \text{power in coil at supply frequency} = \frac{1}{2}\omega\phi_3\phi_1^3[1.9125 + 47.188(\phi_3^2 + (1/2)\phi_1^2)] \sin 3\theta = -P_{e_1}$$

Using equations 12 and 13 yields

$$P_{e_1} = \text{power in coil at the subharmonic frequency } \omega/2\pi = \text{power dissipated in the circuit at the same frequency} = \frac{1}{2} I_1^2 R = \frac{1}{2}(B_1^2 + B_2^2)R \qquad (20)$$

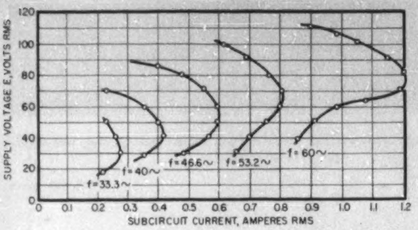


Fig. 14. Experimental synchronizing ranges of third subharmonic oscillations. C=100 μf, R=10 ohms

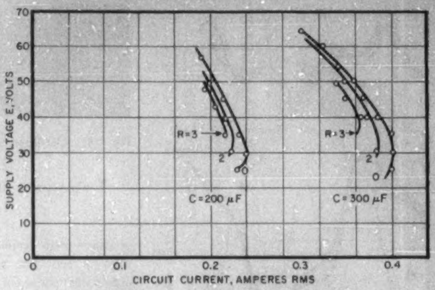


Fig. 15. Experimental synchronizing ranges of fifth subharmonic oscillations

and

$$(1/2)EI_3 \cos \alpha = \frac{1}{2}(I_1^2+I_3^2)R \tag{21}$$

where

$$I_3 = (A_1^2+A_2^2)^{1/2}$$

Again, equations 14 and 15 give

$$Q_{c_1} - \frac{1}{2} I_1^2/\omega C = 0 \tag{22}$$

and

$$Q_{c_3} - \frac{1}{2} I_3^2/3\omega C = \frac{1}{2} EI_3 \sin \alpha \tag{23}$$

where

Q_{c_1} = reactive power in coil at frequency $\omega/2\pi$

$= \frac{1}{2}\omega\phi_1 B_1,$

and

Q_{c_3} = reactive power in coil at frequency $3\omega/2\pi$

$= -\frac{3}{2}\omega\phi_3 A_1$

Eliminating α between equations 21 and 23 gives

$$Q_{c_3} - (1/2)I_3^2/3\omega C = \pm(1/2)[E^2I_3^2 - (I_1^2+I_3^2)^2R^2]^{1/2} \tag{24}$$

From equations 20, 22, and 24 complete information about the third subharmonic oscillation may be graphically obtained, aided generally with auxiliary curves.

Fig. 9 shows the relation between E and $(I_1^2+I_3^2)^{1/2}$ for different values of R and for C=100 μf (microfarads) giving the different ranges of synchronization, E being variable. Lines of constant 3θ are shown in the same figure in dashed lines. It is clear from Fig. 9 that a synchronization curve covers a range of the angle 3θ which decreases as R increases. The relation between E and $(I_1^2+I_3^2)^{1/2}$ for R=0 represents at the same time the two constant 3θ–lines for 3θ=0 and 180 degrees. With the use of equation 17 and aided with auxiliary curves, the regions of stability are obtained; see Fig. 10.

It should be noted here that the discontinuity in the synchronizing ranges for low values of resistances, Figs. 10 and 11, is due to instability. This can be physically attributed to the fact that at such a portion of the curve the amount of detuning is large and this is why it occurs at low values of resistances. The smaller the resistance the larger the variation of natural frequency with amplitude; Fig. 5.

To check the validity of the theoretical predictions regarding the amplitudes and phases of the different components constituting the wave of the sustained oscillations, one of the experimental cases, namely E=62 volts (maximum), C=100 μf, and R=30 ohms, is calculated, giving the following values:

Supply voltage e_s=62 cos ($3\omega t$+58 degrees 46 seconds)
Line current i=0.447 cos $3\omega t$+0.543 cos (ωt+31 degrees 30 seconds)
Flux linkage ϕ=0.217 cos ($3\omega t$+15 degrees) +0.545 cos (ωt+48 degrees 20 seconds)
Voltage across capacitor v_c=14.2 sin $3\omega t$+51.9 sin (ωt+31 degrees 30 seconds)
Voltage across reactor=v=−68.2 sin ($3\omega t$+15 degrees)−57 sin (ωt+48 degrees 20 seconds)

The theoretical waveforms are shown by dashed lines on the corresponding experimental oscillograms, Fig. 12, indicating satisfactory agreement.

Oscillations Just Outside the Locking Range

It was found experimentally that just outside the locking range, the subharmonic oscillations may exist but with slowly varying amplitudes and phases; see the oscillogram of Fig. 13. Just outside the locking range, a change takes place in the instantaneous rate at which energy is supplied from the voltage source to the oscillatory circuit. This results in disturbing the phase balance of the subharmonic oscillation with a subsequent change in frequency. The con-

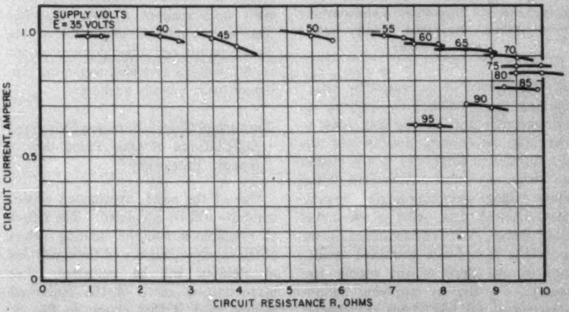


Fig. 16. Experimental synchronizing ranges of second subharmonic oscillations

tribution of the voltage source to the instantaneous rate of energy supplied favors any oscillation of a frequency nearer $n\omega/2\pi$ and thence pulling takes place. The oscillations may therefore be defined as "pulled oscillations." They can be considered as harmonic or subharmonic oscillations but modulated in both amplitude and phase.

In the following analysis, it is assumed that the variation in the amplitude of the flux linkage of the subharmonic oscillation is very small and neglected. This assumption is experimentally satisfied in most cases.

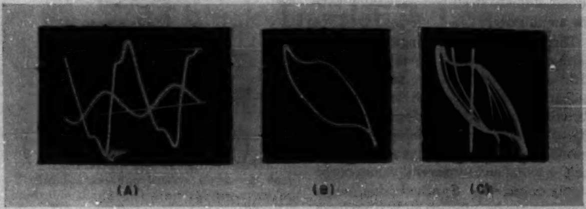


Fig. 17. Lissajous figures of third subharmonic oscillation indicating that subharmonic oscillations should build up in synchronized state. Cartesian waveforms at steady state also indicated

A—Cartesian waveforms B—Steady state C—Building up

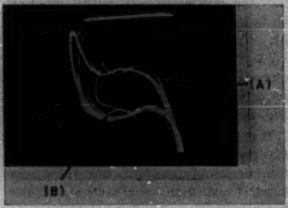


Fig. 18. Lissajous figure indicating that oscillation starts freely and continues for about 2 cycles of natural frequency and then pulls in and continues synchronized until steady state

A—Start
B—Instant of pulling in

Consider the pulled oscillations just outside the third subharmonic locking range and let the flux linkage be given by $\phi=\phi_1 \cos\int(\omega+\theta')dt+\phi_3 \cos 3\omega t$, where θ' stands for $d\theta/dt$ and θ is a variable phase angle; $(\omega+\theta')$ is the instantaneous angular frequency.

Differentiating ϕ once and twice, and substituting in equation 2, after differentiating, yields equation 28 containing

$$\theta', \theta'', \theta'^2$$

$$\begin{matrix}\sin\\ \cos\end{matrix}\int(\omega+\theta')dt$$

$$\begin{matrix}\sin\\ \cos\end{matrix}\,3\theta$$

$$\text{and}\ \begin{matrix}\sin\\ \cos\end{matrix}\,3\omega t$$

where $\theta''=d^2\theta/dt^2$

Usually θ' is very slowly varying compared with ω so that θ'' and θ'^2 may be neglected. At the same time, θ' might be considered constant during an interval covering several cycles of oscillation. Therefore the coefficients of $\cos\int(\omega+\theta')dt$, $\sin\int(\omega+\theta')dt$, $\cos 3\omega t$, and $\sin 3\omega t$ might be separately equated with zero during this interval. Equating coefficient of $\cos\int(\omega+\theta')dt$ with zero, and rearranging, gives

$$d\psi/dt=k_1\,(1+k_2 \sin\psi) \tag{25}$$

where

$$\psi=3\theta-\beta,\ \tan\beta=(1/\omega C_{eq}R_{eq})(D+G)/(D-G)$$
$$k_1=(3/2)\omega[(1/\omega^2C_{eq})(A/\phi_1)-1]$$
$$k_2=(3/2k_1\phi_1)\{[(1/\omega C_{eq})(D+G)]^2+[R_{eq}(D-G)]^2\}^{1/2}$$

A, D, G are given in the Appendix.

Equation 25, although valid at the specified small interval of time, could be satisfactorily considered the general phase-frequency relation at any time t.

From the three relations resulting from equating the coefficients of $\sin 3\omega t$, $\cos 3\omega t$ and $\sin\int(\omega+\theta')dt$ separately with zero together with the solution of equation 25, ϕ_1, ϕ_3, θ, and α may be obtained for given E, C, R, and a certain iron-cored reactor.

Solution of equation 25 is

$$\tan[(3\theta/2)-\beta/2]=-k_2+(1-k_2^2)^{1/2}\tan[(k_1/2)(1-k_2^2)^{1/2}(t+t_0)] \tag{26}$$

where t_0 is a constant of integration.

It is clear from equation 26 that if $|k_2|<1$, θ is periodic and pulled oscillations result. However, if $|k_2|>1$, the tangent of the right-hand side of equation 26 will become a hyperbolic tangent, and θ will go asymptotically toward a constant value and the oscillations will be locked. The range of locking is therefore given by $k_2=\pm1$.

If $|k_2|<1$ (periodic case), then as time t goes on, $[(k_1/2)(1-k_2^2)^{1/2}(t+t_0)]$ passes through $\pi/2$, $3\pi/2$, $5\pi/2$, etc., and the tangent of the right-hand side of equation 26 passes through $+\infty$, $-\infty$, $+\infty$, etc., hence $[(3\theta/2)-(\beta/2)]$ will pass through $\pi/2$, $3\pi/2$, $5\pi/2$, etc. This means that 3θ possesses the same period as $k_1(1-k_2^2)^{1/2}t$. But, while $k_1\,(1-k_2^2)^{1/2}t$ increases uniformly with time, 3θ varies in a nonuniform manner. Therefore

$$\text{average rate of phase variation}=\theta'\,(\text{average})=(k_1/3)(1-k_2^2)^{1/2} \tag{27}$$

The apparent beat frequency is θ'(average), which is observed experimentally, between the oscillation under consideration and another signal of one third the supply frequency.

Very far from locking, k_2 approaches zero and θ' goes asymptotically toward $k_1/3$ and the oscillation will continue freely, but of course damped, with a frequency $[\omega+(k_1/3)]/2\pi$. As the system goes toward locking, the average frequency of oscillation $[=\omega+\theta'(\text{average})]$ will be shifted (pulled) away from the free frequency toward the locked frequency $\omega/2\pi$; see equation 27. This is again a reason for the resulting oscillations being defined as "pulled oscillations."

It is clear from equation 26 that the waveform of the instantaneous phase angle θ depends on k_1, k_2, and β, i.e., it depends on ϕ_1 and ϕ_3 and hence on the resistance R, the capacitance C, and the supply voltage E. It is also clear from equation 27 that the apparent beat frequency depends on ϕ_1 and ϕ_3 and hence on R, C, and E. Near locking, k_2 approaches unity and the period of θ' is very long so that the oscillations appear as if they were nonperiodic.

Experimental Investigation

To test the validity of the foregoing theoretical relations and conclusions regarding the analysis of sustained oscillations occurring in iron-cored reactor oscillatory circuits, as well as to investigate thoroughly the behavior of these oscillations, the following experimental work was carried out.

The magnetization curve passing through the tips of the magnetizing cycles, of the iron-cored reactor used in the experiments, is shown in Fig. 2. Since the natural frequency of the circuit plays a big role in the maintenance of the sustained oscillations, the freely decayed oscillations are recorded on a moving film; see Fig. 3. A 50-cycle timing wave was deliberately recorded on the same oscillogram for comparison. The oscillogram shows clearly that the frequency varies along the wave train; as the amplitude decreases, the frequency decreases. For the sake of comparison, theoretical waveforms of current and capacitor charge are shown in Fig. 4. The waveforms are calculated, representing the magnetization curve by four straight lines. Several other waveforms, for different values of R, are calculated. With these waveforms, the natural frequency against amplitude of oscillation for different values of R is obtained; see Fig. 5.

For the iron-cored reactor used in the experiments, sustained oscilllations of frequencies of one half, one third, one fourth, one fifth, one sixth, and one seventh the supply frequency are obtained. Fig. 11 shows the voltage-current characteristics of the third subharmonic oscillations for different values of capacitance and resistance and for a supply frequency of 50 cycles. For the purpose of comparison the theoretical as well as the experimental curves of the third subharmonic oscillations are shown in Fig. 10, indicating satisfactory agreement.

Fig. 14 shows the voltage-current characteristics for different values of supply frequency, the capacitance and the resistance are of values 100 μf and 10 ohms respectively. Fig. 15 shows the voltage-current characteristics of the fifth subharmonic oscillations for different values of C and R. It is clear from Figs. 11 and 15 that the ranges of synchronization decrease as the resistance increases.

It is also clear that the synchronizing ranges cover a larger current variation for larger capacities than for smaller ones and which is also observed in sustained oscillations of other subharmonic frequencies. These observations may be analyzed as follows.

For possible synchronization, the natural frequency of the circuit should not be far from the required subharmonic frequency; therefore as the capacity increases the effective inductance should be reduced, and hence the current should increase; see Fig. 11. The change of current for a given amount of effective inductance change increases as the current decreases; this is due to the saturation characteristic of the core. Therefore, for a certain amount of detuning between the natural frequency of the circuit and the frequency of the required subharmonic, the required change in current is higher the larger the condenser, agreeing with Figs. 11 and 15. For this reason, it is preferred from a design point of view to obtain sustained oscillations with large values of capacitance.

It should be noted that the synchronizing ranges extend for much larger variations in capacitance and frequency respectively than synchronization with electronic tubes; this is because the natural frequency of the iron-cored oscillatory circuit adjusts itself automatically with the amplitude.

The synchronizing ranges for subharmonics other than the third and the fifth are comparatively small. Even subharmonics have much smaller synchronizing ranges than odd subharmonics; Fig. 16 shows synchronizing ranges of second subharmonic sustained oscillations for $C=100$ μf and for several values of the supply voltage E. Synchronizing ranges of even subharmonics are highly increased by polarizing the core.

Mechanism of Starting Sustained Oscillations

To study the mechanism of generating sustained oscillations (the third subharmonic being considered), the output voltage across the condenser was applied to the X-plate of a cathode-ray oscillograph. A third steady-state subharmonic oscillation obtained by a separate circuit is applied to the Y-plate. Both oscillations on X- and Y-plates are partially filtered with half-section low-pass filters. The Lissajous figure in Fig. 17(C) shows clearly that the oscillation starts and builds up in a synchronized state. The waveforms as well as the Lissajous figure at the steady state are also shown in the same figure. During building-up, energy, being obtained from the supply, must be injected to the oscillation at the right instant and in the right direction. Therefore the oscillation must build up in the synchronized state and remain synchronized in the steady state. However, if starting is aided by an initial high charge on the condenser, the oscillation starts freely and damps until the conditions favor synchronization, and the oscillation is pulled in and continues synchronized until the stationary state is reached; see oscillogram of Fig. 18.

Oscillations Just Outside the Locking Range: Pulled Oscillations

As stated before, these pulled oscillations are actually subharmonic oscillations but modulated mainly in phase. Since the apparent beat frequency is very low; recording on a moving film is a convenient method for studying these oscillations, see oscillogram of Fig. 13. A 50-cycle timing wave was recorded on the same oscillogram for comparison. The apparent beat frequency was very low so that the oscillations in Fig. 13 appear as if they were nonperiodic.

Appendix

$$D_1=\frac{1}{2}\,\omega\phi_3\phi_1^3\left[\frac{3}{4}\,a_3+\frac{15}{16}a_5(2\phi_3^2+\phi_1^2)\right]$$

$$D_2=\frac{5}{32}\,\omega a_5\phi_3\phi_1^5$$

$$D_3=\frac{\omega}{2}\left(\frac{3}{4}\,a_3\phi_1^2\phi_3+\frac{15}{16}\,a_5\phi_1^4\phi_3+\frac{15}{8}\,a_5\phi_1^2\phi_3^3\right)$$

$$D_4=\frac{5}{16}\,\omega a_5\phi_3^2\phi_1^4$$

$$\begin{aligned}D_5=-\frac{1}{2}\,\omega\Bigg[\bigg(&3a_1\phi_3^2+\frac{3}{4}\,a_3\phi_3^4+\frac{9}{2}\,a_3\phi_3^2\phi_1^2+\\&\frac{15}{8}\,a_5\phi_3^6+\frac{45}{4}\,a_5\phi_3^4\phi_1^2+\frac{45}{8}\,a_5\phi_3^2\phi_1^4\bigg)+\\&\bigg(a_1\phi_1^2+\frac{1}{4}\,a_3\phi_1^4+\frac{3}{2}\,a_3\phi_3^2\phi_1^4+\\&\frac{15}{8}\,a_5\phi_3^4\phi_1^2+\frac{15}{4}\,a_5\phi_3^2\phi_1^4+\frac{5}{8}\,a_5\phi_1^6\bigg)\Bigg]\end{aligned}$$

$$\begin{aligned}D_6=-\frac{1}{2}\,\omega\Bigg\{\bigg[&\frac{3}{4}\,a_3\phi_3\phi_1^3+\frac{45}{8}\,a_5\phi_3^3\phi_1^3+\\&\frac{15}{16}\,a_5\phi_3\phi_1^5\bigg]+\bigg[\frac{3}{4}\,a_3\phi_3\phi_1^3+\frac{15}{8}\times\\&\qquad a_5\phi_3^3\phi_1^3+\frac{25}{16}\,a_5\phi_3\phi_1^5\bigg]\Bigg\}\end{aligned}$$

$$\begin{aligned}A_1=(&0.25\phi_3-1.912\phi_3^3-3.82\phi_3\phi_1^2+\\&15.32\phi_3^5+91.92\phi_3^3\phi_1^2+45.96\phi_3\phi_1^4+\\&(-0.6375\phi_1^3+45.96\phi_3^2\phi_1^3+\\&\qquad 7.66\phi_1^5)\cos 3\theta\end{aligned}$$

$$A_2=(0.6375\phi_1^3-15.32\phi_3^2\phi_1^3-7.66\phi_1^5)\sin 3\theta$$

$$\begin{aligned}B_1=(&0.25\phi_1-1.912\phi_1^3-3.82\phi_3^2\phi_1+\\&45.96\phi_3^4\phi_1+15.32\phi_1^5+91.96\phi_3^2\phi_1^3)+\\&(-1.912\phi_3\phi_1^2+45.96\phi_3^3\phi_1^2+\\&\qquad 38.3\phi_3\phi_1^4)\cos 3\theta\end{aligned}$$

$$\begin{aligned}B_2=(&-1.912\phi_3\phi_1^2+45.96\phi_3^3\phi_1^2+\\&\qquad 22.98\phi_3\phi_1^4)\sin 3\theta\end{aligned}$$

$$\begin{aligned}-\phi_1(\omega+\theta')^2&\cos\int(\omega+\theta')dt-\phi_1\theta''\times\\&\sin\int(\omega+\theta')dt-\phi_3(3\omega)^2\cos 3\omega t+\\&R_{eq}\{[(\omega+\theta')(D-G)-3\theta'(D+G)]\\&\sin\ 3\theta\ \cos\ \int(\omega+\theta')dt-[(\omega+\theta')\end{aligned}$$

$$(A+(D+G)\cos 3\theta)-3\theta'(D-G)\times \cos 3\theta]\sin\int(\omega+\theta')dt-3[(\omega+\theta')B-(\omega-\theta')F]\cos 3\theta\sin 3\omega t-3[\omega H+((\omega+\theta')B-(\omega-\theta')F)\sin 3\theta]\cos 3\omega t\}+(1/C_{eq})\{[A+(D+G)\cos 3\theta]\cos\int(\omega+\theta')dt+(D-G)\sin 3\theta\sin\int(\omega+\theta')dt+[H+(B+F)\cos 3\theta]\cos 3\omega t-(B-F)\sin 3\theta\sin 3\omega t\}=-3\omega E\sin(3\omega t+\alpha) \quad (28)$$

where

$$A=a_1\phi_1+(3/4)a_3\phi_1^3+(3/2)a_3\phi_1\phi_3^2+(5/8)a_5\phi_1^5+(15/4)a_5\phi_1^3\phi_3^2+(15/8)a_5\phi_1\phi_3^4$$

$$B=(1/4)a_3\phi_1^3+(5/16)a_5\phi_1^5+(5/4)a_5\phi_1^3\phi_3^2$$

$$D=(3/4)a_3\phi_1^2\phi_3+(5/4)a_5\phi_1\phi_3^4+(15/8)a_5\phi_1^2\phi_3^3$$

$$F=(5/8)a_5\phi_1^2\phi_3^3$$

$$G=(5/16)a_5\phi_1^4\phi_3$$

$$H=a_1\phi_3+(3/2)a_3\phi_1^2\phi_3+(3/4)a_3\phi_3^3+(15/8)a_5\phi_1^4\phi_3+(15/4)a_5\phi_1^2\phi_3^3+(5/8)a_5\phi_3^5$$

References

1. Oscillations in Systems with Nonlinear Reactance, V. L. Hartley. *Bell System Technical Journal*, New York, N. Y., 1936, p. 424.

2. Analysis of Series Capacitor Application Problems, J. W. Butler, C. Concordia. *AIEE Transactions*, vol. 56, Aug. 1937, pp. 975–88.

3. Subharmonics in Circuits Containing Iron-Cored Reactors, Irven Travis, C. N. Weygandt. *Ibid.*, vol. 57, Aug. 1938, pp. 423–31.

4. Beat Theory of Nonlinear Circuits, E. G. Keller. *Journal*, Franklin Institute, Philadelphia, Pa., Sept. 1939, p. 319.

5. The Effect of Initial Conditions on Subharmonic Currents in a Nonlinear Series Circuit, Stephen J. Angello. *AIEE Transactions*, vol. 61, Sept. 1942, pp. 625–27.

6. Negative Resistance Effects in Saturable Reactor Circuits, J. M. Manley, E. Peterson. *Ibid.*, vol. 65, Dec. 1946, pp. 870–81.

7. Forced Oscillations with Nonlinear Restoring Force, Chihiro Hayashi. *Journal of Applied Physics*, New York, N. Y., vol. 24, Feb. 1953, pp. 198–207.

8. Subharmonic Oscillations in Nonlinear Systems, Chihiro Hayashi. *Ibid.*, May 1953, pp. 521–29.

An Electroluminescent Digital Indicator With Elpak Translation Logic

E. A. SACK
ASSOCIATE MEMBER AIEE

Synopsis: In many instances our advancing technology today demands that information be presented to the human observer in digital rather than analog form. A very desirable digital indicator can be achieved if a set of mutually insulated back electrodes or segments are formed on an electroluminescent panel. By applying excitation to selected combinations of these segments, bright numerals or other figures may be formed. The logic required to translate from the input information to the segment code can be accomplished with a number of conventional devices. However, a relatively new switch, known as an Elpak relay, has proved to be well suited to this application.

THE PHENOMENON of electroluminescence is finding a wide range of device application quite apart from its obvious potential as a general source of illumination. This paper will describe an electroluminescent device, known as a digital indicator, which provides a very desirable means for visual presentation of numerical information in digital form. Another electroluminescent device, the electroluminescent-photoconductive relay, or Elpak, has been used to effect the necessary translation for properly exciting the digital indicator. This will also be described.

In recent years, there has been a strong trend toward the presentation of information to the human observer in digital rather than analog form, i.e., the output member of the modern system displays the information as a number or symbol rather than as the displacement of a pointer. The significance of this trend is evidenced by the availability of a growing selection of digital instruments such as digital voltmeters and counters.

A number of disadvantages may be cited for the conventional analog meter. Not the least of these is the fact that many systems, particularly those which utilize digital computation, far surpass the accuracy capabilities of the conventional analog indicator. For example, it would be ridiculous to throttle the potential accuracy of a 6-place counter with an analog output member such as a 1% D'Arsonval meter. Moreover, in a conventional analog instrument only a small portion of the meter face is active in informing the observer of the desired data. For example, assume that the needle on a 4-inch-square panel meter is pointing to the figure 2. In recognizing this reading, the observer is concerned with an area of, perhaps, 1/8 by 1/2 inch on the meter face or about 0.4% of the panel area occupied by the instrument. In a day and age when panel and switchboard space is at a premium, this is very poor utilization of space. Finally, the presence of a multitude of figures and lines on a conventional meter face is confusing to the observer and results in hesitation in reading, interpretation errors, and observer fatigue. The multimeter, with its confusing array of scales and markings is a prime example of how serious this problem can be.

The digital indicator is a distinct step toward overcoming these disadvantages; the accuracy of such a device can be equal to or surpass that required by the system. The digital display to be described occupies no more panel space than is required to "write" down the numeral, and size, and hence the viewing distance; and is completely flexible. Reading errors and fatigue are greatly reduced with the digital presentation because there is no need for the observer to worry about scale factors or to interpolate between scale marks.

Several different types of digital indicators are available. One model utilizes a plurality of individually excited neon lamps spaced behind a mask with numerical cutouts. Another uses painted tabs which are moved, mechanically, into the viewing window. A more modern approach is represented by a device which is essentially a multielectrode neon lamp in which the active electrodes are in the shape of numerals. Another relatively new unit is, in effect, a tiny theater in which the numerals are projected onto a ground glass from the rear by a set of incandescent lamps. Undoubtedly, still others are familiar to the reader. Each

---

Paper 58-150, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 15, 1957; made available for printing November 20, 1957.

E. A. Sack is with the Westinghouse Electric Corporation, Pittsburgh, Pa.

The writer wishes to acknowledge the significant contributions to this work made by several of his associates. R. F. Hollandbeck, G. Kuntz, and J. Tucker designed and constructed the various models described. Dr. R. C. Lyman and Dr. W. M. Kaufman are actively pursuing studies closely related to various aspects of this work and many of their ideas are reflected in the text of the paper.

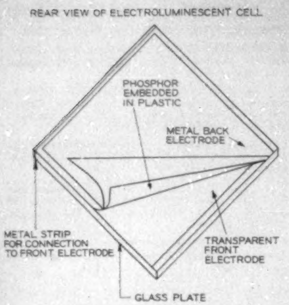

REAR VIEW OF ELECTROLUMINESCENT CELL
PHOSPHOR EMBEDDED IN PLASTIC
METAL BACK ELECTRODE
METAL STRIP FOR CONNECTION TO FRONT ELECTRODE
TRANSPARENT FRONT ELECTRODE
GLASS PLATE

Fig. 1. Westinghouse electroluminescent panel

possesses certain advantages and disadvantages. It is believed that the device to be described is potentially superior to these earlier approaches in a number of important respects.

Electroluminescent Digital Display

Fig. 1 shows a Westinghouse electroluminescent panel. A transparent conducting layer is provided on the back of a glass base plate as the first or front electrode. Next, a plastic-phosphor dielectric is sprayed on the panel. Finally, an aluminum film is evaporated on the dielectric to form a second or back electrode. In effect, the structure is a parallel plate capacitor with a composite phosphor-plastic dielectric.

When a time-varying electric potential is applied to the electrodes, the alternating gradient in the phosphor causes it to luminesce. The theory and technology of intrinsic electroluminescence have been extensively reviewed in the literature and need not be repeated here.[1] Let it suffice to summarize the pertinent terminal characteristics of such a "lamp" as follows:

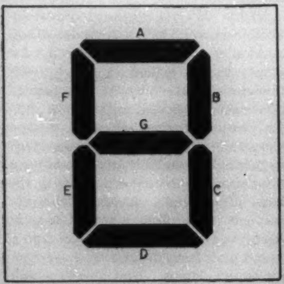


Fig. 2. Digital indicator segment pattern

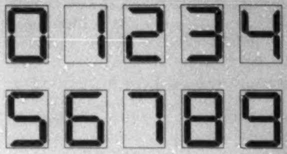


Fig. 3. Numerals formed using the 7-segment pattern

1. The brightness of an electroluminescent cell generally increases with applied voltage and frequency.
2. The color of the light depends upon phosphor composition and frequency.
3. Electrically, an electroluminescent cell looks like a slightly nonlinear capacitor with a power factor on the order of 10% to 20%.
4. Build-up time to full light output is no more than several cycles of the applied power.

The aluminum back electrode need not be a continuous sheet. Instead, the aluminum may be evaporated through a mask to form any desired pattern of insulated electrodes on the back of the panel. When power is applied to any one of these electrodes or segments, only the phosphor under that electrode will luminesce. As a matter of fact, since typically the phosphor-plastic dielectric is only a fraction of a mil thick, the effects of fringing about the electrode are almost entirely negligible.

Contact may be made to the aluminum segments in a variety of ways. For temporary experimental setups, conducting rubber is ideal, since the connection may be established repeatedly without damage to the aluminum film. For finished devices a variety of printed-circuit contact techniques may be utilized.

If the back electrodes are evaporated in a pattern similar to that shown in Fig. 2, any numeral from 0 to 9 may be formed by applying excitation as indicated in Fig. 3 and Table I. Naturally, the transparent conducting electrode is common. Fig. 4 shows a 2-digit indicator in operation.

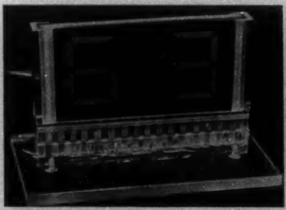


Fig. 4. A 2-digit indicator in operation

365
WESTINGHOUSE

Fig. 5. A 3-digit indicator in a counter

It is interesting to note that various observers report individual preferences in how the seven segments should be shaped to give the most pleasing display. This is not surprising in view of the fact that humans vary widely in the way they shape numerals when writing them. Some like a slanted form while others prefer the vertical shape. However, all observers agree that no difficulty is experienced in recognizing the numbers formed using the 7-segment code.

The shape and size of the numerals on the electroluminescent plate is, of course, completely flexible. More than one number may be put on a single plate, or single numeral plates may be arranged side by side to achieve any number of digits. Other segments may be included to provide polarity marks, decimal points, commas, arrows, vector marks, etc.

Fig. 6. Construction of the 3-digit indicator

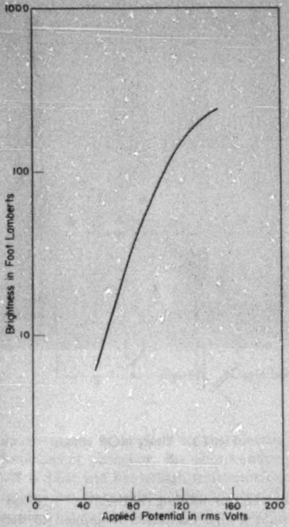


Fig. 7. Brightness of low-voltage electroluminescent panel as a function of applied voltage

Table I. Segment Excitation Program to Form Digits 0 Through 9

	Digit Formed									
Segment	0	1	2	3	4	5	6	7	8	9
A	X*		X	X		X	X	X	X	X
B	X	X	X	X	X			X	X	X
C	X	X		X	X	X	X	X	X	X
D	X		X	X		X	X		X	X
E	X		X				X		X	
F	X				X	X	X		X	X
G			X	X	X	X	X		X	X

* X denotes excitation applied.

An electroluminescent digital indicator for a 3-digit counter is shown in Fig. 5. The construction of this panel is shown in Fig. 6. The electrode segments for the three digits are first evaporated on the electroluminescent plate. A printed-circuit board is prepared with conductors to carry power to the segments through a set of holes drilled to coincide with the segments. The printed-circuit board and electroluminescent panel are then glued together and conducting cement is poured in the holes to establish contact between the printed conductors and the segments. The finished panel may be plugged into a standard printed-circuit base as shown.

There are other fabrication techniques which have been used successfully. By the use of the proper plastic interlayer, the printed circuit may be put right on the electroluminescent panel. The inert plastic layer prevents excitation of the phosphor under the conductors.

Power Requirements

The voltage and frequency needed to power the electroluminescent indicator depend on the brightness desired and the cell construction. Brightness requirements, of course, depend on the ambient lighting conditions. In a low light environment as little as 10 to 30 foot-lamberts might be completely adequate. This is comparable to the highlight brightness at which people tend to operate their television sets for living-room viewing conditions. In surroundings with moderate light levels, such as laboratories, control rooms, etc., 50 to 80 foot-lamberts may be required, while in a high-ambient location, e.g., in direct sunlight, several hundred foot-lamberts may be barely adequate.

Under high-intensity excitation conditions, brightnesses of 2,000 or more foot-lamberts have been attained with plastic-embedded electroluminescent panels.[2] Recent improvements in phosphors and cell construction make it possible to achieve excellent light output with relatively low-voltage excitation. Fig. 7 shows the brightness of a low-voltage Westinghouse electroluminescent panel at a frequency of 10 kc. It may be seen that sufficient brightness for most applications is available with less than 150 volts rms applied to the cell.

At 100 volts, 10 kc, the power dissipated in this type of electroluminescent panel is approximately 50 milliwatts/cm^2 (per square centimeter). For the indicator shown in Fig. 5, the maximum power for each numeral is only 66 milliwatts under typical operating conditions. The capacitance of the low-voltage panel is

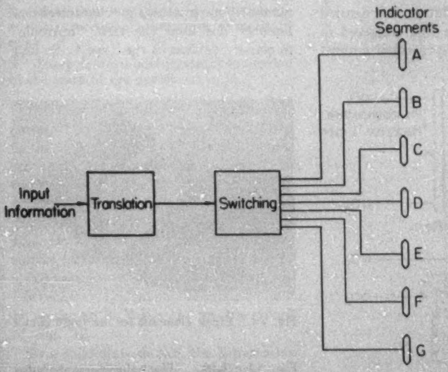


Fig. 8. Translation and switching for electroluminescent indicator

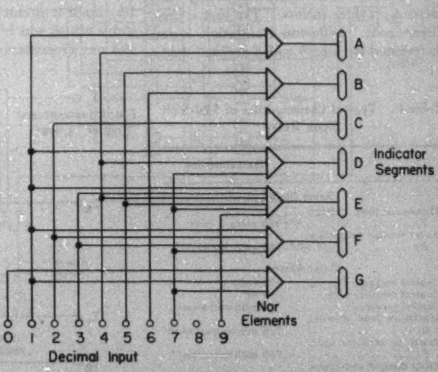


Fig. 9. NOR element translation circuit for the indicator

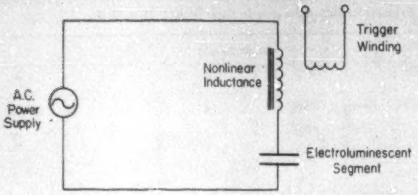
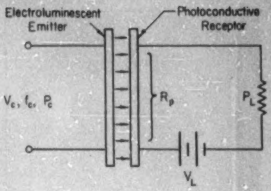


Fig. 10 (left). Ferroresonant switching circuit for the indicator

Fig. 12 (right). Elpak relay: load circuit



approximately 280 micromicrofarads/cm^2 so that the capacitance of each segment, in the display of Fig. 5, is less than 140 micromicrofarads. Under these circumstances the excitation for ten digits or more may be supplied by a small transistor converter.

Switching and Translation

A 2-part process is involved in selecting the proper combinations of segments to form the numeral indicated by the input signal. As shown in Fig. 8, the input information, be it in binary-coded decimal or decimal form must be translated into the 7-segment code. Power must then be switched to the proper segments in accordance with the code signals.

Starting with no power applied to the screen, and switching to turn elements ON, 49 closures are required to form the ten digits. On the other hand, starting with all segments ON and switching proper combinations OFF, only 21 switching operations are involved. The latter case is obviously the more desirable, i.e., it is more efficient to use switching logic which removes, rather than applies excitation to, the segments in order to form the desired digits.

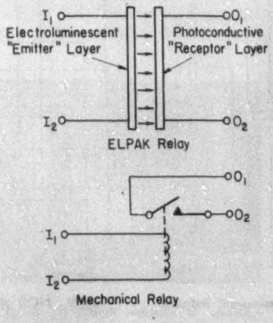
The form of the translation circuit will depend on three factors. The first is the input code. Different translation logic is required for binary-coded-decimal than straight-decimal input. The translation circuitry will also depend on the logic elements to be used. Finally, the form of the logic circuitry will vary with the components used to apply and remove excitation from the segments.

Table II. Typical Characteristics of 120-Volt Elpak Relay

	Characteristics
Relay Open	
Maximum load voltage, V_L	120 volts
Relay contact resistance, R_P	300 megohm-cm^2
Relay Closed	
Control voltage, V_C	120 volts
Control frequency, f_c	5 kc
Control power, P_C	24 milliwatts/cm^2
Maximum load current, i_L	7.5 milliamperes/cm^2
Power to resistive load, P_L	750 milliwatts/cm^2
Relay contact resistance, R_p	2,680 ohm-cm^2

Fig. 9 shows a NOR element translation circuit which may be used when a signal received on one of ten busses (decimal code) is to be displayed as a corresponding digit using the 7-segment code. The NOR element is a component which provides an output only so long as no signal is received on any of its input leads. When a signal appears on one or more of the leads, the output disappears.

Quite often the components available for the logic circuitry, such as small transistors or diodes, cannot handle the voltages which must be applied to and removed from the segments. Then it is necessary to provide output devices which, when driven by the logic signals, will perform the switching operation. Among the elements which may be used for this purpose are vacuum tubes, transistors with step-up transformers, saturable cores and semiconductor devices which exhibit thyratron characteristics. A very satisfactory approach is to make the electroluminescent segment part of a ferroresonant circuit, as shown in Fig. 10. Such a circuit exhibits two distinct stable states which may be triggered in and out of existence by a small current in a trigger winding on the core.[3] In one state a high excitation is applied to the segment, while in the other the segment is dark.



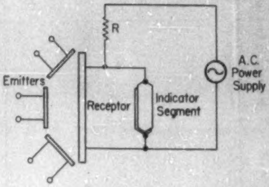


Fig. 13. Elpak NOR circuit

The next section describes still another switching device which is capable of performing both the translation and the switching functions. In many ways this element is ideally suited to perform these operations for the electroluminescent indicator.

Elpak Logic

If an electroluminescent cell and a photoconductor are associated as shown in Fig. 11, there results a switching device, here called an Elpak relay, which has a number of interesting characteristics. Loebner has used the term "optronic" to classify devices of this type.[4] So long

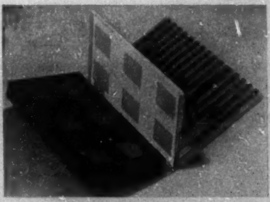


Fig. 14. Elpak elements for the logic circuit

Fig. 11 (left). Electroluminescent-photoconductive Elpak relay

Fig. 15. "Logic box" for the Elpak plates

as no power is applied to the electroluminescent phosphor, the photoconductor is dark and has a high resistance between its terminals. When power is applied to the electroluminescent "emitter," however, the illuminated photoconductive "receptor" assumes a resistance which may be five orders of magnitude below its dark state.

In many ways the Elpak relay is closely analogous to the mechanical (magnetic) relay, certainly far more so than either the vacuum tube or the transistor:

1. There is excellent electrical isolation between the control and load terminals.
2. The load terminals are bilateral. (No bias is needed to switch dual polarity current.)
3. More than one receptor may be arranged to view a single emitter, thus simulating a multipole relay.
4. More than one emitter may illuminate a single receptor to simulate a multicoil relay.
5. There is an inherent visual indication of the state of the device.

The control power and allowable load power of an Elpak relay are proportional to the active area of the electroluminescent and photoconductive surfaces. In the case where the two layers are congruent and directly adjacent, the terminal characteristics may be specified for a unit area of the relay. For a 120-volt zinc sulfide electroluminescent emitter facing a 120-volt photoconductive cadmium sulfide receptor, typical characteristics might be as shown in Table II; see the circuit of Fig. 12.

As a solid-state device, the Elpak relay has advantages of life, reliability, and ruggedness over the mechanical relay. At present, however, it has a disadvantage which limits its application. Sensitive photoconductors, such as cadmium sulfide, are slow to respond to variations in illumination. Depending upon the light levels available, from 0.1 to 1 second may be required for the photoconductor to change from one extreme of conductivity to the other.

The translation circuit for an electroluminescent digital indicator has been built with the use of the Elpak relay. This seemed to be appropriate for a number of reasons. First, the idea of using electroluminescent logic with the electroluminescent indicator was quite intriguing. Second, the power-handling capabilities of available photoconductors are such as to obviate the need for a separate control stage following the logic. Third, impedance levels and excitation requirements of the indicator segments and Elpak relays are excellently matched and, finally, the speed of response of the photoconductor appeared to be adequate for this application.

The Elpak logic circuit which was constructed used Elpak NOR elements to translate from decimal to the 7-segment code with the arrangement shown in Fig. 9. An Elpak NOR circuit is shown in Fig. 13. So long as no power is applied to any of the electroluminescent emitters, excitation is applied to the display segment. When a signal is received on one or more

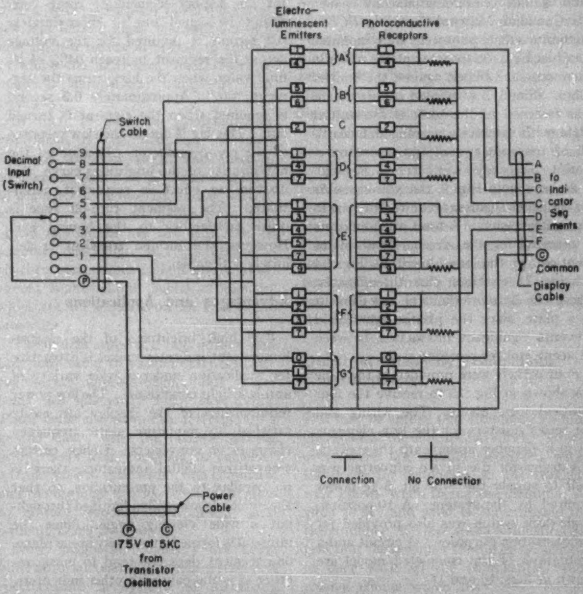


Fig. 16. Circuit of electroluminescent translation device

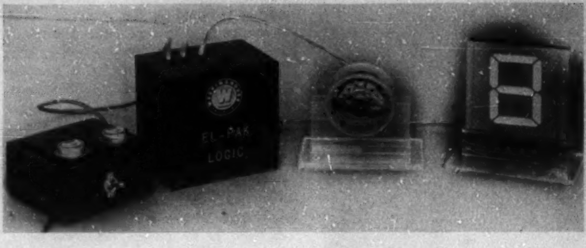


Fig. 17. Demonstration indicator with Elpak translation logic

of the emitters, the photoconductive receptor removes the output excitation.

In the model which was constructed, it was convenient for geometrical reasons to use a separate photocell for each emitter and then connect all of the photocells for a particular NOR element in parallel. This resulted in a receptor with a power-handling capacity which was many times that actually required.

The electroluminescent NOR elements were fabricated in plate form as shown in Fig. 14. Six emitter electrodes were evaporated on the back of a 2- by 4-inch zinc sulfide electroluminescent panel. Six Canadian Marconi Company *PC1-04* cadmium sulfide photocells were mounted in a black plastic receptor plate, which in turn was sandwiched against the emitter plate. Finally, a printed-circuit board was fastened to the back of the emitter plate with conductors arranged to carry all of the emitter and receptor connections to a special socket in the "logic box."

As seen from Fig. 9, the NOR elements for the seven segments require from one to six emitters each. A total of 21 emitters is required for the screen. Four plates identical to the one shown in Fig. 14 provide 21 emitters plus three spares. There can be more than one NOR element in a plate, since the plastic light shield prevents significant interaction between adjacent emitter-receptor pairs.

Four sockets were provided in the logic box shown in Fig. 15 to receive the four identical logic plates. The wiring and the series resistors for the NOR elements are in a chamber underneath the sockets.

A transistor d-c to a-c converter was built to supply the 175-volt 5-kc power required by the system. A 10-position, single-deck switch was also provided for demonstration purposes. A circuit and a photograph of the completed model are shown in Figs. 16 and 17.

The performance of one Elpak NOR element and associated display segment is, of course, typical of all of them. Tests show that so long as the emitter is off, 165 volts of the 175 volts applied appear across the segment. Under these circumstances, the brightness of the segment is on the order of 80-foot-lamberts. When power is applied to the emitter, the potential across the display segment drops to about 16 volts. With this excitation, the segment is "out" (less than 0.01 foot-lambert).

With the Elpak relay, it takes longer for the display segment to come "on" than to be turned "out." Approximately 0.2 second is required for the voltage across the segment to reach 64% of its final value when the logic turns the segment "on." Approximately 0.3 second is required when the element is turned out. This lag is due to the slow response of the photoconductor. Because of the nonlinearity of the brightness curve, this uneven lag effect is somewhat accentuated. The segment turn-off time is quite satisfactory to the human eye. However, the turn-on transient is definitely noticeable.

Advantages and Applications

The high brightness of the electroluminescent indicator makes it attractive for application under a wide variety of ambient-light conditions. The low power requirements of the display are easily satisfied by miniature static frequency changers or converters. Unlike certain competitive digital indicators, there is no parallax to the presentation, so that the numerals are easily identified throughout a wide viewing angle. Since the numeral is formed completely in one plane, one segment does not tend to mask another as is the case with other indicators. The possibility of varying the color of the electroluminescent numeral adds another degree of freedom to the display.

The size of an electroluminescent panel is limited only by manufacturing convenience. One-foot-square cells are already available. Numerals over one foot high may be made by using more than one panel, since there need be no dead area around the edges of the cell.

On the other end of the scale, numerals may be made whose height is no more than a fraction of an inch. As noted earlier, a display on an electroluminescent panel may have very fine resolution.

The potential utilizations of an indicator of this type are too numerous to list in entirety. There are applications in automation, process supervision, testing and quality control, computer readout, annunciation, and score display. Since the form of the numeral and auxiliary markings is easily varied, the electroluminescent indicator may be tailored to match the application.

The techniques described in the foregoing may be carried over into precoded graphic displays for power system switching, pipe-line-flow monitoring, and industrial process supervision. Here, in particular, the outstanding color flexibility of electroluminescent lighting will be a great asset.

Conclusions

The electroluminescent digital indicator provides a numeral display with low parallax, high brightness, excellent readability, low power requirements, variable color, and wide dimensional flexibility. The logic and control which are required to translate information from a particular code to the 7-segment code may make use of a wide selection of conventional devices. The Elpak relay provides a desirable means for accomplishing both the translation and the control in a single step. The Elpak logic circuit which was built to translate from the decimal to the indicator code has proved to be very satisfactory.

References

1. ELECTROLUMINESCENCE AND RELATED TOPICS, G. Destriau, H. F. Ivey. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 12, Dec. 1955, pp. 1911–40.

2. RECENT ADVANCES IN LUMINESCENCE, H. F. Ivey. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 13, 1957.

3. THE FERRORESONANT TRIGGER PAIR: ANALYSIS AND DESIGN, Charles F. Spitzer. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 407–16.

4. OPTO-ELECTRONIC DEVICES AND NETWORKS, E. E. Loebner. *Proceedings*, Institute of Radio Engineers, vol. 43, no. 12, Dec. 1955, pp. 1897–1906.

Introducing Young Engineers to the Appreciation of Magnetic Amplifier Problems

L. A. FINZI
MEMBER AIEE

YOUNG ENGINEERS entering a field of electrical development and design probably are equipped with some initial knowledge of that area, but may lack a balanced perspective of the relative significance of phenomena and facts and methods. In fact, they may have altogether immature notions of the ways by which engineering problems are solved. This is particularly true if their college training has been of a dreary "learning-by-the-book" variety, where procedures and "formulas" may have been derived or accepted without adequate effort to appreciate their physical significance, limitations and implications, and where so-called problems really may not have extended beyond disguised pretexts for algebraic or arithmetical manipulations within the rigid frame of those formulas.

In many fields the young apprentice is assisted by dependable elements of experience gathered by others and expressed, for instance, by company files and collections of test results on earlier products, and by the advice offered by senior colleagues on the job. This body of knowledge should not be considered as unassailable and final, but certainly it is useful in bringing the problems at hand within the grasp of the beginner. The situation is more difficult in newer fields, and in particular in the rapidly moving field of nonlinear ferromagnetic devices, with which this paper is concerned.

Magnetic amplifiers are deceptively simple devices in their structures and in their essential circuitries; it is not difficult to acquire some notion of their general operation in a qualitative way, and the design of the components is initiated with rather elementary considerations that are not different from those of small transformer design. But a few steps beyond this initial stage the real difficulties appear.

The ways of thinking and the techniques introduced in the analysis of different systems seem dubious to the beginner, and vary inconsistently from case to case; the most important material properties appear elusive to an extent unsuspected at first. Though the technical literature is now emerging from the confused state of earlier years and good treatises are becoming available, the unconventional character of this nonlinear field is conducive to misinterpretations, and the immature reader may not be immediately able to make good use of these sources. Moreover, in proceeding to a reduction to practice in a design with quantitative specifications of performance or in a semiquantitative development of new and untried ideas, many effects of a most unclear nature are likely to appear, for which the individual must resort to empirical approaches, using his own ingenuity.

In fact, it may be admitted that empiricism is the mainstay of most magnetic amplifier work, and there is nothing disgraceful in this admission since all branches of engineering and all natural sciences have evolved in this way; empirical approaches are not antithetic to scientific reasoning. But a fruitful empiricism presupposes insight and critical, almost sophisticated, habits of thought which are not acquired on the spur of the moment.

These difficulties are not unique and they may occur to various extents in any engineering apprenticeship, but newer fields are inherently more mystifying. These fields also seem characterized by a greater haste, with the restless shifting of engineers from company to company or from task to task within one company, and the beginner is often required to grow to mature responsibilities quickly and without much assistance.

Now, magnetic amplifiers are being built quite successfully, meeting stringent specifications in systems of increasing complexity with a wealth of clever new ideas and disclosures concerning the use of the basic components, and all this is done to a large extent by young engineers not far from their college years and equipped perhaps with no more than some tables of core and wire sizes. This proves that many college graduates are endowed with intelligence, ingenuity, and remarkable endurance. On the other hand, the ordinary process of amplifier development and design usually demands a great effort, is costly, and involves a great many engineering man-hours.

Here the question arises whether effective forms of training may be devised for the swift growth of competent designers. The aims are to economize on some part of the time that the young engineers must spend searching their own ways and learning from their own mistakes, and further to rescue individuals who in the process may coast into spiritless technical activities of an indiscriminate hit-and-miss variety.

A schooling of this sort should not be conceived as a teaching of special technologies for special tasks. Factual information and manuals of design, etc., are most useful in competent hands but they do not in themselves point the way to attaining competence. The latter relies on the acquisition of keen mental habits in the detection and interpretation of the amplifier phenomena; in the considerate choice of meaningful methods, analytical and experimental; and in the selection of whatever simplifying assumption may seem expedient for quick and significant answers.

It remains to be seen whether anything can be done practically and in a short time to achieve some of these lofty aims. The considerations that follow express some tentative ideas on this matter. Apologies are made for the belaboring of some issues that may seem self-evident to many readers, but the educational process is made difficult in this field by the extreme ease with which misconceptions arise in the beginner's mind.

Quest for Simplicity; Simplifying Assumptions

It may be agreed that simplicity should characterize the educational effort. However, there are pitfalls associated with the desire for simplicity, as any teacher well knows when facing the dilemma of what to say and what to leave unsaid to the student. No one would seriously suggest that the beginner should be presented at once with all the complexities and elusive effects which may well emerge significantly in his future work. But neither is a starting point obtained in

---

Paper 58-177, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing November 13, 1957.

L. A. Finzi is with the Carnegie Institute of Technology, Pittsburgh, Pa.

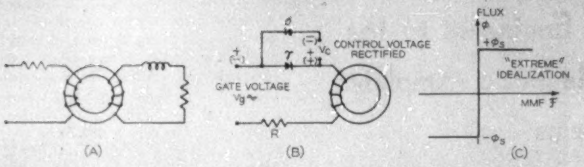


Fig. 1. A—A simple reference circuit. B—A voltage-controlled half-wave amplifier. C—The extreme idealization of core properties

explanations which explain hardly anything constructively.

For instance, the idea that the magnetic amplifier is "sort of a reactance" is misleading and also finds little use in any further description of amplifier circuits. The action of the amplifier in the load circuit is by no means that of inductance; also a coil wound on a core of a square-loop material really does not possess any one of the properties inherent to the concept of inductance. The terms "saturable reactor" or "variable impedance" are used in the slang of the art, but this is not a reason for building on such usages.

The beginner, who by definition knows no better, is bound to be literal-minded and should not be exposed needlessly to malapropisms and far-fetched analogies. One is aware of the usefulness of certain linearizations of core properties in certain aspects of certain problems, but this is not a matter for generalizations out of context. The beginner's background in magnetic circuits may be already anchored on sophomoric treatment of the subject in college in which there was unwillingness to depart from linear or linearized procedures (with hysteresis loops injected, none too clearly, as a matter of curiosity). Therefore need may be felt early in the beginner's training to shock him into the ways of square-loop thinking which are more pertinent in many problems of modern magnetic-amplifier applications.

Forms of simplicity that are educationally questionable are also forms in which a comparatively complicated system is made simple by the precipitate introduction of extreme idealizations. The series-connected saturable reactor with resistive loads and zero control-circuit resistance, and with ideal cores of infinite and zero permeabilities, is an example of these approaches. Such treatments quite skillfully describe certain interesting patterns of behavior, and their place in the art should not be belittled. But strong reservations are voiced here in principle against the postulation of assumptions and against the immediate use of reasonings by limiting cases in the educational approach without careful and reasoned justification.

The beginner is not in the position to appreciate what the assumptions basically stand for and can only accept passively the methods. Soon enough, however, he may recognize that his interests as a designer are in, for instance, the behavior of saturable reactors at low signal levels or in their operation as even-harmonic modulators, where the assumption of unimpeded flow of harmonics is self-defeating; or in extension to doublers and center-taps whose finite gains cannot forever be complacently attributed to leaky rectifiers, and so on. A search for better procedures may be difficult at this point if exceedingly simplified ways of thinking have been already ingrained in the mind of the uncritical beginner. Probably, because he has been exposed to a vague treatment of unspecified magnetizing currents, interactions, and harmonics, he may infer, with frustration, that the "theories" have little bearing on his practical problems; he may find it difficult to reach meaningful conclusions by himself, and plunge into brute-force attempts to meet his immediate assignments.

Moreover, exceedingly simplified treatments tailored to fit a particular device do not lead to the coherent understanding of the many other systems or devices grouped under the comprehensive heading of "magnetic amplifiers," and the experience that one may have gathered in the earlier study of a certain system remains unrelated to newer tasks. This is a strange result, since in all cases the same basic components are found utilized with essentially similar mechanisms of operation.

The approach suggested in the following is one in which simplicity is sought in a different manner. A physical system is considered first which is comparatively simple in itself, i.e., a core with some windings. A fascinating wealth of meaningful problems may be found even at this level in applications of practical interest, while a later transition to more complex multicore system may coherently evolve.

No simplifying assumptions are postulated a priori for this; rather the properties of actual components (cores, rectifiers, etc.) are considered, as known from the recent technical literature and manufacturers' data and as recognized by experimentation, and the essential meanings of the fundamental laws expressed in the system equations are examined. It is only at this point that a search for idealizations and simplifications may be initiated profitably for any specific problem at hand.

In any problem of engineering for which no pat solution is known in advance (in which case, it would not actually be a problem), the choice of approximate simplifying assumptions is, more or less, a matter of calculated risks. Whenever one is required to replace an actual situation with an idealized (and thus different) one, the possibility of later "agonizing reappraisals" is implicit. This point may not have been stressed in college and this may be one reason for the hasty statements of disillusioned engineers to the effect that theories are of no use in the solution of their practical problems. Magnetic-amplifier problems require a most inquisitive attitude in these choices. Students should be encouraged by all means to use their ingenuity to devise and try simplified approaches, extreme as these may be, but some discrimination must be fostered throughout the process.

Organization of Course Content

General Features; Voltage Reset

The implications of Kirchhoff's voltage law as applied to one toroidal winding with resistances or other components in series may be reviewed first. It is suggested that the habit is formed of writing down the circuit equations carefully, simply because this is the most convenient and compact way to express the physical laws. Even when it is found that the solution of these equations may be cumbersome or impossible, their judicious consideration and reconsideration at all steps constitutes valuable guidance in partial predictions, in the interpretation of empirical findings, and in extrapolations by similitude.

Concepts of flux and flux linkages may be reviewed to emphasize the volt-time integral idea and to point out the principle of nonrectification for circuits supplied from periodic voltage sources of zero cyclic average. Stray fluxes and partial flux linkages and winding capacities may be mentioned without loss of simplicity because these concepts are within the student's grasp. No particular restrictions are needed, so far, concerning the core properties. In view of the highly nonlinear problems to be tackled later, inductance concepts L and M may well be omitted. Essentially, inductances are not fundamental physical quantities, but rather man-made definitions, introduced as an abbreviation of a sort, convenient but not indispensable even in the treatment of linear circuits.

Square-loop cores are considered next, and systems with cores of "gradually varying differential permeability" are postponed to a later time as distracting in the introduction of novel concepts of more immediate interest. The properties of square-loop cores are expressed well enough at this point by the major loops, static and dynamic for sinusoidal fluxes, and by dynamic minor loops exhibited in catalogs of core manufacturers. Some of these loops as well as the peculiar wave forms of magnetizing currents may be displayed on an oscilloscope with elementary techniques. (Direct experimentation should be an integral part of the training in this field in which the oscilloscope remains the primary tool.)

A comparison with similar loops of other materials (e.g., hot-rolled steel) reveals three peculiar features of square-loop materials:

1. Sharply defined saturation flux densities: $\pm B_s$.
2. Tendency of the core flux to remain at a set value upon removal of the magnetomotive force.
3. Comparatively small field intensities in flux excursions over the unsaturated range.

Feature 1 leads to the definition of the core volt-time area and of the normal excitation. A simple application is found, for instance, in the reference circuit of Milnes,[1] in its basic form; the influence of changes of line voltage and frequency may be investigated and the usefulness of the secondary network may be recognized; see Fig. 1(A). The inverter of Royer[2] may be considered next; the less than perfect "squareness of the top of the loop" is a useful factor in the operation of this inverter, while, on the other hand, this effect may well be ignored in many other instances.

Feature 2 is utilized most evidently in Pittman's time-delay device[3] and in many other transistorized devices. The thyratronlike action of a half-wave amplifier with resistive load may be investigated next in terms of the preset flux level ϕ_p. The well-known voltage-reset half-wave circuit of Fig. 1(B) may be used to show one particular way in which ϕ_p can be made dependent on a signal v_c. Hence transfer characteristics and gain concepts are presented, with careful and specific definitions to avoid ambiguities that are not unusual.

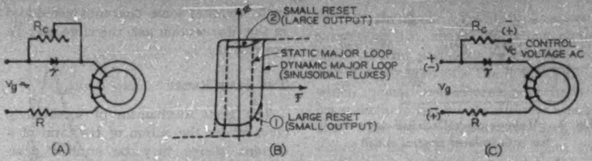


Fig. 2. A—A resistance-controlled half-wave amplifier. B—Utilizing the information of core catalogs. C—An instance of control by nonrectified alternating voltages

Cases of sinusoidal gate voltages and rectified sinusoidal signals are followed by conditions of more general waveforms of signal and gate voltages, larger or smaller than normal. Some fundamental formulas of design, for the core and the gate winding, may be presented at this stage, which are common to most magnetic amplifiers and some comments may be given on wire tables, winding factors, and shuttle limitations.

Other types of loads, resistive-inductive, with or without discharging rectifiers, with d-c back electromotive forces, etc., may be considered next, with the recognition that the concept of firing leads to load-circuit analyses similar to those of thyratron situations, while the time of extinction of gate-winding currents affects the control, as it determines the time at which reset can be initiated by the signal. The coincidence counters, the ring counters, the digital components of Ramey,[4] and the multiplier of Mathias, et al.,[5] may be examined also.

Feature 3 has been used in all previous instances only to note that certain resistive voltage drops may be neglected in the voltage equation. (Some numerical examples of the relative order of magnitude of the various terms of the voltage equation should be given.) That is, the extreme 3-zone idealization of Fig. 1(C) relating core flux ϕ to core mmf (magnetomotive force) is accepted for preliminary purposes, though this matter may deserve reconsideration, depending on the accuracy required in the various cases.

The resistance-controlled half-wave circuit of Fig. 2(A) may be considered next, as a further instance in which the voltage equations still describe the operation. But here the controlling agent is the voltage drop of resetting currents through the resistor R_c; hence some notice must be taken of these currents. The information concerning core mmf must be reexamined, and on this occasion some additional comment is in order concerning dynamic widenings of the major loops for sinusoidal fluxes at the various frequencies. The student familiar with eddy currents will accept the idea that larger mmf may be needed at higher values of rate of change of fluxes, perhaps to compensate for the reaction of induced eddy currents, although the shapes and widenings of the dynamic loops of the catalog may indicate to him that more is involved than the eddy currents of his sophomoric experience.

A prediction of the transfer characteristic, output versus R_c, is possible, if it is noted that large resets, minor loop 1 in Fig. 2(B), are characterized by rather large values of $d\phi/dt$, in dynamic processes which are similar to those of sinusoidal voltage of rated frequency, with comparatively large resetting currents throughout, while small resets (minor loop 2) are associated with small resetting currents; for intermediate outputs the reset currents have somewhat intermediate values. This is rough reasoning which, however, is not without practical value; the essential phenomena have been expressed crudely but have not been ignored. The student will understand how core-testing procedures could be sought if more accurate predictions were important to him and how certain idealizations of core properties (e.g., the rectangular remanator concept or the empirical equation $\mathcal{F} = \pm\mathcal{F}_c + K\,d\phi/dt$) could be used to expedite his work. A similar approach is required in the example of Fig. 2(C).

The 2-winding amplifier circuit of Fig. 3 is essentially a pretext to review transformer concepts. The beginner, under the cursory college treatment of short-circuit and no-load tests, vector diagrams, etc., pertinent to power transformer techniques, may have missed transformer concepts of more basic interest. Hence there may be need to write the differential equations associated with the various windings, introducing mmf with clear sign conventions and dot notations. The point should be made that the instantaneous terminal quantities describing the core operation are its flux ϕ, the flux rate (or voltage per turn) $d\phi/dt$, and the mmf $\mathcal{F} = \Sigma Ni$ resultant from the instantaneous

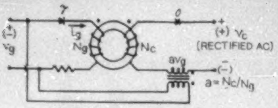


Fig. 3. Voltage-controlled half-wave amplifier with isolated control circuit

currents in all windings; techniques of reductions of secondary quantities to primary in terms of turns ratios $a = N_1/N_2$ and ratios squared, etc., result then directly.

The N^2/R concept for the two windings should be emphasized, associating it with their copper volumes; hence another concept of design is introduced which is most important in questions of optimum design of all amplifiers. (No harm is done if formulas for self-inductances and mutual inductances and Q's are recalled at this point in design.)

This is also the time, dispelling any confusion, to point out that voltage gains arising from ratios of transformation are different from the gains inherent in amplification.

The peculiar features of a saturable-core transformer are displayed in the system of Fig. 4, which provides short high-energy pulses for stitch-welding purposes upon periodical reversal of the d-c line voltage. A tentative design of the system components may be attempted, for given pulse specifications, under the assumption of negligible secondary inductance and with some idealization of the weld resistance. The simultaneous use of the concepts of volt-time integrals and of current-time integrals is interesting. As a separate problem, the oscillatory discharge may be investigated that takes place if the secondary inductance is large (assuming, for simplicity, that the switch is opened as soon as the core reaches saturation).

Multicore combinations of the basic voltage-reset half-wave circuit may be studied next. Fallacies of indiscriminate "building-block" ideas may be detected by examining the poor operation of a-c-output 2-core systems obtained by connecting two of the blocks back-to-back; here the coupling through the load branch makes it difficult or even impossible to obtain a reasonable operation; i.e., the transfer characteristics are not comparable to those of the basic half-wave circuit. In this connection, Maine's papers on single-phase and polyphase combination of half-wave components[6] are worth-while reading. (The 2-core half-wave reversible-output amplifier of Lufcy[7] and others similar to this may also be examined, although a more appropriate treatment really requires some concepts introduced in the next section and therefore may be postponed.)

THE MMF RESET

The reset mechanism of square-loop cores under the action of the mmf of a current source may be studied next. This mechanism appears in the control of various types of magnetic amplifiers and in the switching of memory cores and transfluxors. Oscillographic tests of reset core-voltage versus time are performed easily (Fig. 4) conveying more specific ideas of dynamic behavior (i.e., the three terminal quantities $\mathfrak{F}$, ϕ, and $d\phi/dt$ are interrelated, at least for the particular process of flux reversal from pronounced saturation). Industrial methods of core testing and matching may be described.

Use of oscillographic tests of core reset voltages has been made by Critchlow, et al.,[8] to predict transfer characteristics of half-wave and 2-core self-saturating amplifiers controlled by constant-current sources. The processes may be examined that take place in gate-limited intervals (i.e., in those intervals of the reverse half-cycle during which the self-saturating rectifier becomes conducting and the core reset is limited by the gate voltage) and in constant-mmf intervals (in which the rectifier blocks). The rather low limiter voltages recognized in the doubler after firing, once more warn against unwarranted building-block notions, and the influence of gate-winding resistances is seen.

The mechanism of operation of these self-saturating circuits over the negative branch of their transfer characteristic is also easily described. This leads to consideration of the operation around minimum output and to the introduction of reasonable speculations concerning those triggering phenomena which are particularly evident at gate voltages lower than normal. As these phenomena are associated with flux swings initiating from less than pronounced saturation, the expedient of auxiliary pulses of bias mmf suggested by Morgan,[9] becomes understandable; also some reasons are seen why rounded-top materials (such as 48-alloy) are less critical in terms of triggering.

The biased-rectifier-amplifier (BRA) of Morgan[10] may now be examined in some of its simplest configurations to emphasize the importance of gate-limiter voltages and in discussing more closely the behavior of rectifiers and their threshold voltages.

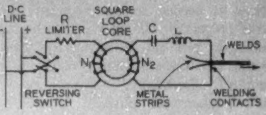


Fig. 4. Stitch-welding system

The system of Fig. 6 is suggested as a possible means to obtain two voltage pulses e_2 and e_3 in sequence, upon application of a current pulse I_1. Kirchhoff's laws for magnetic circuits implemented by some dynamic core information, perhaps simplified to the analytical form $H = H_c + K\,d\phi/dt$, yield a straightforward algebraic treatment resulting in significant predictions, while the **B-H** idealization with infinite and zero permeabilities, or its modification into a rectangular loop of fixed width, would not be fruitful. (It is interesting to point out here that an eddy-currentlike dynamic behavior appears also in ceramic materials in which eddy currents do not flow.)

Empirical core-reset functions, relating $\mathfrak{F}$, ϕ, and $d\phi/dt$ through flux swings from pronounced saturation, have been suggested by Feth,[11] from certain methods of dynamic testing on grain-oriented Ni-Fe alloys. These functions are intended to supply the missing link in the system of equations (Kirchhoff's voltage equations for the various meshes and mmf expressions for the various cores) describing the reset of all amplifiers examined so far; i.e., complete analytical treatments of those systems are made possible. It is not suggested here that these treatments are indispensable in questions of practical design; the point is that the student, aware by now of the general problem in its broad terms, may revert to simplified treatments, with some mature appreciation of what he is doing and why. In particular, he will be able to rise above the question as to whether the magnetic amplifier is a voltage-controlled or a current-controlled device and, in his work, will choose whatever idealization of the controlling action seems most pertinent.

The amplifier circuit of Fig. 7 is an instance in which the choice of one or the other of these two ways of thinking is prescribed by the operating specifications. In this amplifier the gating operation is based on the use of supply voltages considerably larger than twice the normal voltage of one core. Hence, in the absence of signal, both cores race simultaneously in each half-cycle from an initial saturation towards the opposite saturation, firing simultaneously before the end of the half-cycle, with zero output. The function of the signal is to unbalance the times at which the two cores saturate; output is obtained by transformer action in each half-cycle during the interval in which one core is already saturated while the other is not.

In a fast-response amplifier described by Hill and Fingerett,[12] the signal is the rectified voltage of an a-c transducer. The smallness of the mmf requirements of unsaturated cores justifies a straightforward use of the voltage equations to describe the operation, neglecting voltage drops of magnetizing currents, i.e., using the "extreme" core idealization.

On the other hand, in a low-level application analyzed by Frost-Smith,[13] the signal is a constant current I_c obtained, for instance, from a thermocouple or a photocell through a choke or a parallel inductance-capacitance combination. Here the extreme core idealization, or this together with the assumption of a rectangular major loop of fixed width, would lead to absurd conclusions, since the difference between the mmf of the two cores $\mathfrak{F}_1 = N_g i_g + N_c I_c$ and $\mathfrak{F}_2 = N_g i_g - N_c I_c$ prescribes the difference in the firing angles. Milnes' 3-zone representation may be used in this problem; more modern core-reset functions or their idealization in the form of a parallelogram fitting a suitable major loop (e.g., for sinusoidal flux swings between $\pm\phi_s$, at a frequency increased so as to obtain appropriate values of $d\phi/dt$ at the various ϕ's) may be preferable in a compromise between accuracy and timing of these studies.

It may be interesting to examine also the final version of this amplifier involving the use of two rectifiers λ' and λ'', connected back to back in parallel in the output circuit. This is an instance in which a symmetrical nonlinear resistance-like device is used to ensure balanced output, in remarkable similarity to some versions of Morgan's BRA single-core push-pull circuits.

THE 2-CORE SATURABLE REACTOR AND RELATED CIRCUITS

Self-saturating doublers and center taps, etc., with low control impedance may be considered next, either directly in terms of their circuit equations or following the classic approach which derives them from the 2-core simple saturable reactor.

The saturable reactor with zero control impedance may be treated, starting from the opposite end, i.e., from the case of infinite control impedance, reducing then the control impedance to a finite resistance $R_c > 0$, and ultimately arriving at the limiting case $R_c = 0$. This is suggested to prevent misunderstandings, detected in practice, concerning the meaning and use of the assumption $R_c = 0$.

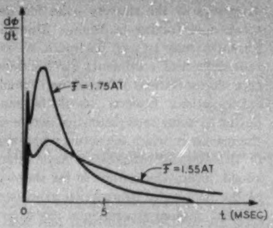


Fig. 5. Core-reset voltages upon application of constant reversing mmf $\mathfrak{F}$ to saturated core

As a "warm-up" for all this the waveform of the current i_g may be investigated in the simple system of Fig. 8(A). With some insight, the steady state may be described directly, but students, in this and many similar instances, may find it difficult to "jump into" the steady state directly if no starting information is supplied concerning the instantaneous values of the variables at some time in the cycle.

Ramey's philosophy,[14] of arriving at the steady state from a transient initiated from specified initial conditions, finds here a particularly simple application, i.e., at some arbitrary time, the switch S is closed and the ensuing processes are followed. (To avoid distracting issues of dynamic core properties etc., a numerical instance may be chosen in this problem in which $N_c I_c$ is substantially larger than any mmf requirements of the unsaturated core.)

The 2-core system of Fig. 8(B) is next considered. With the h-circuit open, this is the d-c transformer for metering purposes which may be treated either by the transient approach (which, however, is not as simple here as in the earlier case) or proceeding from the case of zero output circuit resistance to cases of finite output resistance as is done by Storm in chapter 8 of his book.[15]

Already, it is here seen that the extreme core idealization dictates the relation between instantaneous (and thus also half-cyclic) values of the ampere turns of control and gate circuit, if no simultaneous saturation of the two cores occurs at any time in the cycle. Volt-time area considerations (i.e., the use of voltage equations) remain necessary to examine the core flux, to determine the limits beyond which simultaneous saturation of the two cores disrupts the d-c transformer operation. The even-harmonic voltages induced in the circuit h, when open, are easily determined, and the operation of even-harmonic modulators into systems of infinite input impedance is understood.

The operation is modified when the h-circuit is closed. Simple manipulations prove the equivalence of this system to that of Fig. 8(C) where $R_c = R_h$ (if, for simplicity $N_h = N_c$) and $E_c = R_c I_c$. The control current i_c is then $i_c = I_c + i_h$. The study of this new system (i.e., the general series saturable reactor, with $R_c > 0$) is straightforward in terms of the voltage equations and of the core mmf expressions if the extreme idealization of core properties is accepted (and only resistive loads are considered).

Simple results are obtained if a rectangular gate voltage is assumed with a d-c control voltage E_c or if sinusoidal full-wave rectified signals in phase with a sinusoidal gate voltage are considered. The practical case of sinusoidal gate voltages and d-c control voltages inherently leads to more cumbersome solutions, exhaustively treated by Barker.[16]

At this point, the students recognize that solutions simplify conspicuously if E_c and R_c happen to be reasonably small with respect to $V_g N_c / N_g$ and $R_g (N_c / N_g)^2$ and thus they may appreciate the classic zero-control-resistance treatment of Storm in its meaning and usefulness, with a healthy understanding of the more general situation from which this limiting case has been approached.

Extrinsic feedback up to 100% may now be introduced, with discussion of algebraic equivalence of the latter circuit and of the doubler, and with mention of the different points of view in the choice of what should be considered as the basic amplifier. Perfect rectifiers may be assumed at first, to avoid the confusion of issues; leakage effects are introduced most easily at a later time.

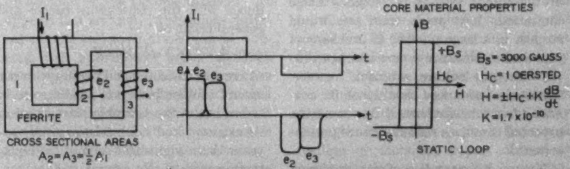


Fig. 6. Generation of pulses with multiaperture core

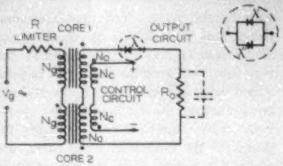


Fig. 7. Amplifier with even-harmonic output

The increasing need for more realistic core representations appears soon in this process. One will not say that the doubler with perfect rectifiers "theoretically has infinite gains, but practically this isn't so," recognizing rather that infinite-gain notions arise from an absurd extension of the use of simplifying assumptions that were acceptable perhaps in other situations for certain modest aims. Conceptual improvements upon the results of the extreme idealizations may be shown by the use of the 3-zone representation of Milnes,[1] which postulates a large but finite permeability for the unsaturated cores. The finite gains of the 100% feedback saturable reactor are justified in this way and the gradual transition from the controlled-current-source character of the saturable reactor to the voltage-output character of the doubler is also seen clearly.

This representation, of course, does not express the memory feature of square-loop cores, as it implies $\phi=0$ when $\mathfrak{F}=0$, but memory is not a first-order feature in the operation of these devices. In a better adherence to the measurable properties of modern cores, Storm's representation may be used as introduced by him in chapter 15 of his book.[16] His assumption, $\mathfrak{F}=0$, when $d\phi/dt=0$ for the unsaturated core, may be modified with the speculations on dynamic and static major loops of Jackson, et al.[17] Tentative criteria are thus obtained concerning the half-cyclic mmf of control.

The main stumbling block for more accurate predictions is recognized in our present inadequacies to express the processes that take place in the time interval in the half-cycle while the bottom of the minor loop is traversed. These limitations hurt more than one would suspect in a large number of problems of practical significance even with regard to the innocent-looking saturable reactor. These circumstances should not be concealed from the students if they are to be expected to make a reasoned use of present criteria.

Now, a different line of attack is suggested, using experimental transfer characteristics as the starting point for the solution of further problems. This information may be used backward, to obtain numerical coefficients for fictitious core representations of further use. Such extrapolations, however, should be supported by some semiqualitative notion of the manner in which the actual properties of the physical components have intervened, instant by instant in the shaping of the half-cyclic information expressed by that original experimental characteristic.

Experimental transfer characteristics are used more directly and quite simply in a large number of problems of design, e.g., to predict effects of degrading or upgrading of rectifiers or of negative or positive feedbacks of other kinds. Again, awareness of the nature of the instantaneous phenomena, as detected by oscilloscopic observation, is needed to prevent misuses of these simple criteria.

Properly measured transfer characteristics supply also the information needed for predictions of transient response. Outputs transients resulting from step changes of signal voltage, signal resistance, loading or line voltage, or frequency are related directly to corresponding changes of the preset flux value ϕ_p throughout the transient. These changes are expressed in the differential equation of the control circuit, introducing half-cyclic integrals as time variables.

It happens that the steady-state transfer characteristic, relating the gross half-cyclic voltage output V_0 (inclusive of the voltage drops in the gate-winding resistances) to the control voltage E_c, offers also a welcome means of expressing $\Delta\phi_p/\Delta I_2$ in the equation, since $\Delta\phi_p/\Delta I_c$ is proportional to the slope $\Delta V_0/\Delta E_c$ of the transfer characteristic, at any value of the half-cyclic V_0. This consideration applies for all magnetic amplifiers of this kind, from the simple saturable reactor to bistables, and opens the way for the integration of the first-order differential equation.

When the static gross voltage-gain $\Delta V_0/\Delta E_c$ is practically constant over the range of outputs considered, the resulting equation is one with constant coefficients and the amplifier time constant τ is introduced with the remarkably simple and general result of

$$\tau=\Delta V_0/\Delta E_c\times N_c/N_g\times 1/2f$$

where τ is in seconds and f is in cycles per second. When both gate windings are in series, i.e., in the saturable reactor and in its extrinsic feedback variations, $2N_g$ appears in the right-hand side of the equation instead of N_g. As is well known, obvious modifications of the general result arise when other circuits, of bias, etc., are present to further couple the two cores.

Every transfer characteristic of the slope $\Delta V_0/\Delta E_c$ applies for steady-state operation, after the step change of the variable that gave origin to the transient. The latter may be known directly or may be obtained from other available characteristics, or it may be predicted from core properties; the inadequacies of the latter procedures, of course, affect also the corresponding prediction of transient response.

The same results are obtained for the time constant of the amplifier proper, when feeding into rectified inductive-resistive loads with discharge paths. For inductive-resistive a-c output loadings, a transfer characteristic obtained from rectified terminal voltage measurements does not supply direct information on the half-cyclic flux swings; predictions of over-all transient behavior may be obtained still from direct steady-state measurements of these flux swings at various I_c's under pertinent loading conditions.

The simplicity of this approach and the results are obscured if that unfortunate concept, the figure of merit, is introduced prematurely in the transient problem. This is true because the concept is one step more remote from the essential mechanism of all magnetic amplifiers (saturable reactor included), i.e., the half-cyclic output voltage is the quantity truly controlled by the absorption of a part of the total half-cyclic volt-time integral of the supply voltage in the way of half-cyclic swings of core flux. Of course, the student will have understood by now that magnetic amplifiers are power amplifiers and that power gains and figures of merit are quantities of foremost practical interest. But this is not a reason for giving untimely emphasis to the latter concept in the transient problem.

So-called transportation lags may be discussed if it is recognized that reasonings in terms of half-cyclic averages inherently do not allow for short-time resolutions. Methods of measuring the transient response may be described.

When half-cyclic values are considered as the time variables of the problem, difference-equation methods suggest themselves; some information on the essence of these methods should be given the student (e.g., to enable him to understand some of the recent literature) considering them as a way of thinking which is appropriate and elegant.

Presently, much work is being done to describe the dynamic behavior of the magnetic amplifier, not only as a system component but also within itself, in the

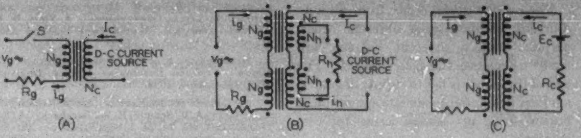


Fig. 8. Introducing the series-connected saturable reactor with finite control resistance $R_c > 0$

language of feedback theory. Expressing a recognized pattern of physical behavior in a different language is often beneficial as this can lead to different points of view and thus also to different and newer applications. However, there is a danger in this: the students, carried away by the relative ease of block diagram and transfer function techniques, should not lose sight of the actual physical phenomena only partially expressed in these treatments.

The transient treatment does not account for delaying effects of instantaneous currents that may be induced to flow at various times in various paths, resulting in a delay of over-all response, with well-known differences of response in building-up and building-down transients, especially in the doubler, and with limitations of the meaningful extension of that treatment to fast transients. While the treatment of these effects are most difficult, their semiqualitative recognition (e.g., in terms of Lenz's law) will assist the student in his later empirical work to appreciate, for instance, the influence of gate-winding resistances in these transients.

Also gains and time constants obtained independently for two amplifiers may fail to give reliable clues for a 2-stage system resulting from their combination. Evidently this is because the instantaneous constraints appearing in the interconnection are not expressed in the half-cyclic average information previously mentioned. This recognition at least opens the way to more pertinent experimentation in empirical developments.

Push-pull and polyphase combinations and self-balancing amplifiers, etc., remain to be examined. Students, commonly familiar with similar techniques in vacuum-tube electronics will recognize to some extent the basic ideas of circuitry. The difficulties met before in the attempts to describe the operation of conventional single-ended amplifiers multiply rapidly in these situations, but it is not within the scope of the training course to supply quick answers to complex questions which are as yet largely unsolved. The semiqualitative consideration of design found in the books of Geyger,[18] Storm, and Milnes, are a good starting point for a reduction to practice. The designer is on his own from there.

To conclude the course, some brief survey may be given of the methods used in the study of amplifiers with cores of gradually varying differential permeability, as presented years ago, e.g., by Johnson[19] and by Beaumariage, et al.[20] A single-valued analytical approximation of the normal magnetization curve and the assumptions of zero control resistance and zero load resistance are used; the concept of firing does not occur in these treatments and the influence of loads is expressed by load lines of sinusoidal a-c theory. These earlier approaches, dictated by the properties of the core materials then most readily available, are rather unsatisfactory in view of the assumptions. But improvements are most difficult analytically and this explanis the place these methods still hold in some European and Russian literature.

The student may gratefully recognize at this point that the strange features of square-loop cores that may have shocked him at first are really most beneficial, performance considerations aside, as leading to more satisfactory and interesting ways of thinking.

Closing Remarks

The program presented is not intended to teach special technologies. The semiempirical state of the art and the many as yet unanswered questions appearing in any development of magnetic amplifiers suggest that it may be necessary, first, to clarify the pertinent physical laws and phenomena of these systems in relation to the essential properties of the basic components. Some subtle issues should not be avoided, which conceivably one would prefer to ignore, but which unfortunately remain of primary concern to anyone interested in quantitative results of design.

Comparatively simple physical instances may be introduced in a coherent sequence, so as to stress a few concepts and methods at the time, with proper emphasis on the appraisal of the simplifying approximations that may or may not be accepted in the various instances. It appears that a variety of seemingly different circuits and applications can be covered with a consistent logic, (although "coverage" should not be considered the essential aim), and that lively and critical habits of thought can be developed which are indispensable for a competent use of the literature and later empirical work.

For this, lectures to a passive audience should be avoided; information passively acquired can seldom be put to use; rather they may result in a dangerous acceptance of ill-digested half-truths when novel and unconventional concepts are unfolded. Many of the instances may be presented as problems, with the assistance of a few restrained hints, leaving the students to search for suitable approaches with the help of some simple experimentations as needed, and then co-ordinating their findings in a more general perspective. In other instances, the students may be invited to establish critical comparisons of of the content of technical papers describing devices unrelated perhaps from the point of view of applications but embodying some essential common feature.

A dozen problems, distributed over about as many weeks, and solved with the support of weekly half-day laboratory sessions and 2-hour class meetings have proved sufficient to cover most of the area described in a recent graduate course at the Carnegie Institute of Technology. The same general approach has been tried also in industry; dealing with groups of engineers, initially not familiar with the magnetic amplifier, but otherwise mature and conversant with questions of development and the design of electric devices, the scholastic "home problem" approach was felt less necessary, but it was still preferred to lead the audience into discussion, rather than "lecturing" it. In either case the results seemed quite good as evidenced by the wealth of original thought, with germs of ideas of the kind that may evolve into technical papers or patent disclosures and worth-while products.

It may be added that students approaching this peculiar field with open mind, considering magnetic amplifiers as electromagnetic systems that obey the general laws of electromagnetism with particular twists dictated by the special properties of the components, did not have serious conceptual difficulties at any time. This was not the case, initially, for students who had previous misconceptions and biased points of view, easily arising in this field, e.g., from the casual reading of technical papers chosen at random or from unwarranted generalizations as a result of brief experience.

The outline given is far from unique; different instances in a different order, perhaps with a different balance, are likely to achieve the same results. The aim of this paper is not to supply dogmatic answers to a difficult educational problem but to bring up certain issues for discussion.

References

1. Transductors and Magnetic Amplifiers (book), **A. G. Milnes.** MacMillan and Company, Ltd., London, England, 1957.

2. A Switching Transistor D-C to A-C Converter, **G. H. Royer.** *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

3. A High-Frequency Static State Delay Device Utilizing Transistors, **G. F. Pittman, Jr.** *Ibid.*, Mar. 1955, pp. 54–58.

4. The Single-Core Magnetic Amplifier as a Computer Element, **R. A. Ramey.** *Ibid.*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 442–46.

5. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, **L. A. Finzi, R. A. Mathias.** *Ibid.*, vol. 72, pt. I, Sept. 1953, pp. 455–61.

6. High-Speed Magnetic Amplifiers, **A. E. Maine.** *Electronic Engineering*, London, England, May, 1954; Nov. 1954.

7. An Improved Magnetic Servo Amplifier, **C. W. Lufcy, A. E. Schmid, P. W. Barnhart.** *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 281–89.

8. Dynamic Core Behavior and Magnetic Amplifier Performance, **L. A. Finzi, D. L. Critchlow.** *Ibid.*, vol. 76, pt. I, May 1957, pp. 229–40.

9. Reaction of Bistable B-H Loops on Magnetic Amplifiers, **R. E. Morgan.** *AIEE CP 57-776* (copies available on request).

10. The Biased Rectifier Amplifier—A Pulse Magnetic Amplifier, **R. E. Morgan.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 584–90.

11. Core-Reset Functions in Magnetic Amplifier Analysis, Part I, **G. C. Feth.** *Ibid.*, vol. 77, pt. I, 1958 (*Paper no. 57-95*).

12. Fast Response Magnetic Servo-Amplifier, **F. Hill, J. A. Fingerett.** *Electronics*, New York, N. Y., Oct. 1954.

13. The Study of a Magnetic Inverter, **E. H. Frost-Smith.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. II, 1953, p. 362.

14. On the Mechanics of Magnetic Amplifier Operation, **Robert A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

15. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955.

16. The Series Magnetic Amplifier, **R. C. Barker.** *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 819–31.

17. The Operation of Magnetic Amplifiers with Various Types of Load, Part II, **L. A. Finzi, R. R. Jackson.** *Ibid.*, vol. 73, pt. I, July 1954, pp. 279–88.

18. Magnetic-Amplifier Circuits (book), **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957.

19. Universal Curves for D-C Controllable Reactors, **Walter C. Johnson, B. C. Merrell, R. E. Alley, Jr.** *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 31–40.

20. Analytical Determination of Characteristics of Magnetic Amplifiers with Feedback, **D. W. Ver Planck, L. A. Finzi, D. C. Beaumariage.** *Ibid.*, pp. 565–70.

Automatic Control of Electrical Precipitation Rectifiers

H. E. VAN HOESEN
NONMEMBER AIEE

H. J. WHITE
MEMBER AIEE

H. J. HALL
ASSOCIATE MEMBER AIEE

THE PRIMARY objectives of an automatic rectifier control in electrical precipitation are to increase long-term collection efficiency, eliminate need for operating personnel, and insure maximum utilization of equipment under the widely variable process and operating conditions found in most applications. A review of the literature[1–3] indicates that a number of automatic control schemes have been devised for precipitator control. Most of these systems never received widespread usage because of inherent limitations. Some required expensive, complicated mechanisms which were difficult to maintain or adjust. Others were unable to provide the correct control range, response or reaction to meet the variety of operating conditions encountered. More recently, improved equipment based on electronic techniques has been developed and is now in successful commercial use.[4]

Experience with automatic rectifier controls for electric precipitators since 1949 has shown the desirability of incorporating voltage, current, and spark-rate as control parameters. Extremely rapid response is essential to cope with sudden changes in precipitator electrical conditions which commonly occur in practice. This paper describes a new automatic control system which meets these requirements and which is broadly applicable to industrial electrical precipitation.

Electrical Characteristics and Control Requirements of Precipitators

Before describing the new control system, it is well to review electrical precipitator operation and the control requirements. The critical dependence of collection efficiency on the operating peak voltage and sparking rate as illustrated in Figs. 1 and 2, is characteristic of electric precipitators. On account of the steep voltage-current characteristic of the corona discharge, small changes in voltage can produce large changes in current and power. This is especially significant in the region of the sparking potential where the corona power increases at a rate at least four or five times faster than the peak voltage. High voltage and high corona power promote high collection efficiency. On the other hand, excessive voltage causes a rapidly increasing sparking rate which reduces the efficiency. Obviously, there exists a balance between these opposing factors where the gains obtained by increasing the voltage are offset by the losses resulting from sparking. For our purposes, it is important to differentiate between a spark and a power arc. A spark, illustrated in the precipitator voltage oscillogram of Fig. 3, is a random, transient voltage breakdown of short duration, considerably less than 1/10 of a voltage cycle of the power frequency. By contrast, a power arc is a continuous heavy-current discharge which lasts from a few cycles to hundreds of cycles. Whereas a spark has only a slight effect on precipitator collection, a power arc is disastrous to performance and should be avoided by proper circuit design.

In practice, electrical characteristics of precipitators vary widely not only between installations, but also change with time on any given installation. Plant process variations and electrode dust deposits affect voltage, current, and sparking characteristics. All precipitator operation can be classified under three general cases.

1. Current limited: The precipitator may have low impedance characterized by heavy current, low voltage, and little or no sparking. The appropriate control action in this case is to provide maximum current flow, limited only by rectifier capacity.

Paper 58-229, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 25, 1957.

H. E. Van Hoesen, H. J. White, and H. J. Hall are with Research-Cottrell, Inc., Bound Brook, N. J.

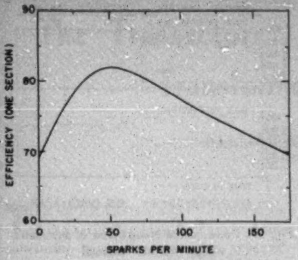


Fig. 1. Per-cent efficiency as a function of average sparking rate for a typical fly-ash precipitator

2. Voltage limited: The precipitator may have relatively high impedance, with low current, high voltage, and little or no sparking. Here the best action is to raise the precipitator voltage as high as possible, consistent with rectifier ratings.

3. Spark limited: Under this condition, precipitator power input is limited by sparking below the current and voltage limits of the rectifier. Optimum collection is achieved by maintaining a reasonably constant spark-rate, typically 100 sparks per minute per rectifier set.

The New Automatic Control System

The new automatic control system is based on the three general cases of precipitator operation outlined in the foregoing. Design requirements include dynamic, wide-range control with fast transient response and high reliability. A simplified block diagram of this system is shown in Fig. 4. Two feedback control loops are used: one for current and one for spark-rate. These control loops operate independently or co-operatively according to the instantaneous control requirements. The voltage level control provides an upward bias which tends to raise the rectifier transformer voltage to its rated value. Error signals from the two feedback loops modify the voltage raise function to establish system equilibrium at a dynamic control point.

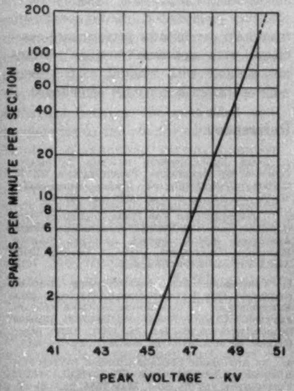


Fig. 2. Variation in sparking rate with operating peak voltage for a representative fly-ash precipitator

A saturable-core reactor is used as the control element to regulate power input to the high-voltage transformer and rectifier. Control of this reactor is accomplished through a grid-controlled thyratron driver which performs the function of a stable high-gain d-c power amplifier. The three control signals, voltage, current, and spark-rate, are injected through the master comparator to the driver. This comparator functions as a mixer which provides an output signal equal to the sum of three input signals.

Current control is accomplished through the current signal generator which converts an a-c signal to an error signal proportional to the deviation from the current control point. Primary current is sensed by the voltage drop across a small resistor in the primary circuit. The current signal can be taken from the secondary circuit if desired.

The spark-rate signal generator converts spark signal pulses to a d-c error signal proportional to the deviation from the desired spark-rate. A simplified schematic circuit of the spark-rate control is shown in Fig. 5. Spark transients are sensed across high-voltage transformer primary winding. These pulses are separated from the a-c component by a spark detecting network which passes only spark pulses. These pulses fire switch tube *V*1 in the pulser circuit, composed of elements *V*1, *L*1, *C*1, and *R*1. Each time *V*1 fires, it discharges condenser *C*1 momentarily. The following charging current for *C*1 produces a voltage pulse across *R*1. This voltage pulse is integrated by the network composed of *R*2 and *C*2 to obtain a voltage proportional to the average spark-rate. The integrated voltage V_A appears across *R*3 as the spark signal voltage. Time constants *R*1 *C*1, *R*2 *C*2, and *R*3 *C*2 must be properly selected to obtain proper integration and control action.

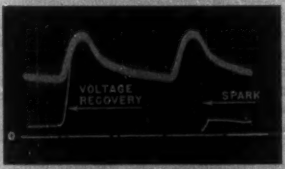


Fig. 3. Precipitator voltage oscillogram for half-wave rectifier connection illustrating rapid recovery after sparking recurrent sweep

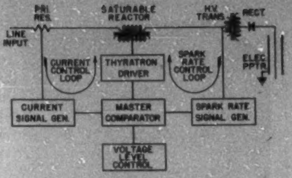


Fig. 4. Block diagram for automatic rectifier control system

To obtain an error signal it is necessary to compare the spark signal voltage with a reference preset to the desired spark-rate. This is accomplished in the comparator circuit composed of *V*2, *R*3, *R*4, and *R*5. The reference voltage V_B governing spark-rate is preset by potentiometer *R*4. Note that reference voltage V_B is in opposition to spark signal V_A. Error output voltage V_C appears across *R*5 only when the signal voltage V_A exceeds the reference voltage V_B. Under this condition, tube *V*2 conducts and an error voltage proportional to the deviation, appears across *R*5. This error voltage V_C modifies the raise bias on the thyratron driver to maintain a fixed spark-rate level.

The current control system operates in a similar manner. In essence, an a-c signal voltage proportional to primary current is rectified, filtered, and compared with a preset reference value to obtain a current error signal.

The new automatic control unit is

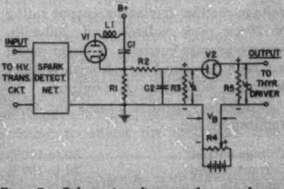


Fig. 5. Schematic diagram for spark-rate signal generator

V_A—Spark signal
V_B—Spark reference
V_C—Error signal

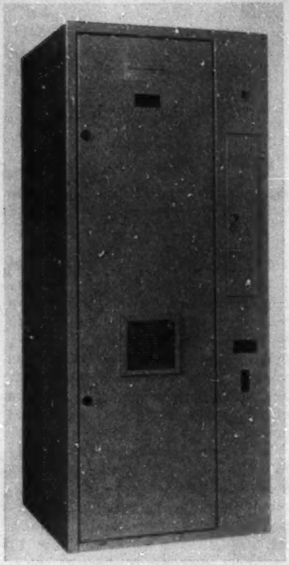


Fig. 6. New automatic control assembly for 230-milliampere rectifier

shown in Fig. 6. Any number of these units can be controlled from a single remote "on-off" switch, an arrangement which is desirable in some installations. Individual service breakers are provided to de-energize each control and its associated high-voltage rectifier. Three factory preset adjustments, voltage level, current level, and spark rate, are provided on each unit. These adjustments may be easily reset in the field if necessary to meet unusual conditions.

This control has been developed primarily for use with 60 cycles per second power systems. The model illustrated is designed for use with a high-voltage rectifier having an output rating of 70-kv peak voltage and 230-milliampere average current. A larger model has been developed for use with rectifier sets up to 500-milliampere output.

Satisfactory control ranges for most applications are:

Voltage output	50 to 100%
Current output	20 to 100%
Spark-rate	50 to 500 per min.

Voltage recovery after a precipitator spark is achieved in one cycle. Transient response time of the current and voltage control is fast enough to insure complete control under all operating conditions. Even for abnormal conditions, such as internal short circuits in the precipitator, the control will automatically limit current at the preset maximum rating, without trip-out or other disturbance. However, a signal system is incorporated to indicate these abnormal conditions. Over-all power conversion efficiency of the complete rectifier system, including all fixed losses, is 70 to 75% at rated output.

Application and Performance

The new automatic control system can be applied advantageously to most industrial precipitator applications such as cleaning of boiler, metallurgical, cement kiln, and chemical process gases. Practically all industrial furnace and process gases are subject to wide variations in gas flow, temperature, humidity, dust concentration, dust resistivity, and other factors which influence precipitator operation. The automatic control tends to compensate for these variations, thereby maintaining optimum precipitator performance.

Field tests on these new control units have been in progress since early 1956 on fly ash and metallurgical applications. The original experimental model has provided continuous trouble-free operation since first installed and has become a permanent part of the plant equipment. Standardized production and commercial application of the new control have been in effect since January 1957. Control levels on voltage, current, and spark rate have been successfully preset at the factory and operation has been entirely satisfactory to date.

The field tests have shown that improved precipitator energization and higher collection efficiencies associated with higher corona power are achieved with this automatic control. Fig. 7 shows a reproduction of recording voltmeter and ammeter charts made on one rectifier set on a fly-ash precipitator to compare manual and automatic control. Because of changes in boiler operation and coal burned, the manual control was normally operated at a moderate power level in order to prevent over-current trip-outs of the rectifier. In contrast, the automatic control operated at substantially higher power because of its ability to rapidly and continuously correct rectifier settings to meet changing load conditions. The average power input to the precipitator from this rectifier set was 5.6 kw with the manual control and 9.3 kw with the automatic control, an increase of 65%. Subsequent periodic measurements have shown this power increase typical.

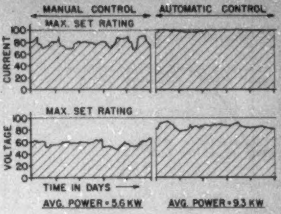


Fig. 7. Performance comparison of automatic versus manual control

Experience on industrial installations has indicated this new control has intrinsic value extending beyond precipitator energization for which it was designed. In certain instances, it is possible to correlate automatic control action with plant process variables. Thus the precipitator control action can be applied as a control parameter for the process variable to provide over-all automatic system control. This not only improves precipitating conditions, but may also increase the uniformity and quality of the product produced.

Conclusion

In conclusion, the automatic rectifier control system described provides optimum precipitator energization under the fluctuating load conditions normally encountered in practice. It is not a substitute for a properly designed high-efficiency precipitator. But it does insure maximum precipitator performance associated with optimum rectifier settings not possible with manual control even with operators in constant attendance.

References

1. Some Developments in the Automatic Control of Electrical Precipitators, H. E. Corbitt. *General Electric Review*, Schenectady, N. Y., vol. 41, no. 4, Apr. 1938, pp. 187–90.

2. Torrlikriktare For Hoga Spanningar (Dry Rectifiers for High-Voltage Applications), Algot Arvidsson. *ASEA Journal*, Allmänna Svenska Elektriska Aktiebolaget, Vasteras, Sweden, 1948, ASEA Reg. 4909 DK621.314.63.

3. Verfahren Zur Selbsttatigen Dauerregelung Der Betriebshochstspannung Von Elektrischen Gasreinigungs-Oder-Prufapparaten. *German Patent no. 644,756*, Siemens Schukertwerke A. G., Jan. 24, 1930. (Also other German patents involving modifications thereof.)

4. An Automatic Voltage Control System for Electrical Precipitators, H. J. Hall. *AIEE Transactions*, vol. 73, pt. I, May 1954, pp. 124–27.

5. Automatic Voltage Control of Electrostatic Precipitator with Saturable Reactor, H. J. Hall. *U.S. Patent no. 2,742,104*, Apr. 17, 1956.

The Transactor: A Self-Saturated Transformer—Part I

ALLEN B. ROSENSTEIN
MEMBER AIEE

THE OPERATION and characteristics of the saturable transformer[1] and the saturating magnetic reference type of transformer[2] have been presented in past articles. The advantages and shortcomings of the self-saturating magnetic amplifier circuits have been explored at length. It is the purpose of this paper to start with the elementary concept of a magnetic switch and develop a magnetically controlled self-saturated transformer. The magnetic system developed possesses many of the desirable operational characteristics of both the saturable and saturating transformers. An adjustable a-c or d-c output is obtained which is also voltage-stabilized. At the same time, the self-saturating transformer retains the gain and response characteristics of the self-saturating magnetic amplifier, although a number of the disadvantages of the magnetic amplifier are eliminated.

Magnetic-Amplifier Shortcomings

The shortcomings of the conventional magnetic amplifier inherently arise from the fact that the saturable reactor (or reactors) is basically connected in series with the load. The reactor acts as a magnetic gate which subtracts a predetermined number of volt-seconds from each half-cycle of a-c source voltage, allowing the remainder to appear across the load. The output thus becomes the difference between the controlled volt-seconds set in the reactor core and what is often a randomly varying quantity, namely, the a-c line voltage. Variations in line voltage become a direct cause of drift and/or noise in the amplifier output. Because of the series form of connection, the magnetizing current of the reactors must flow through the load, giving a quiescent load current. As this quiescent current is often a function of the a-c supply voltage, interstage coupling can thus also present a practical problem. Because of the finite magnetizing current of its reactors, it becomes difficult or impossible for the magnetic amplifier to work into an open circuit or very-high-resistance loads. This sometimes becomes a nuisance for, should the load resistance tend to vary widely, a substantial bleeder resistance is usually required to restrict the variation of output voltage.

Fundamental Magnetic-Switch Operation

The fundamental mechanism of the magnetic amplifier is usually presented with the assumption of a perfect magnetic core. Fig. 1, the "perfect" core, possesses a 2-zone **B-H** characteristic offering infinite permeability when the net ampere-turns impressed upon the core are zero, and zero permeability when the net ampere-turns do not equal zero. With the assumption of perfect cores, the saturable reactors may now be considered as 2-position magnetic switches with a built-in volt-second memory. In their capacity as switches, the reactors in one position (net ampere-turns ≠0) will cause all reactor winding to be short-circuited. In the second position or mode II (net ampere-turns = 0), the reactors appear as perfect transformers, transforming the impedance of the control circuits into the power or gate circuit. In the "game" of magnetic-amplifier circuitry, it becomes our objective to arrange control windings in such a way that negligible induced currents flow in the control windings during this second mode. The control winding are so connected that the voltages induced in them are either self-cancelling, blocked by a high impedance, or bucked out by auxiliary sources. In any event, with negligible induced currents flowing in them, the control windings of the magnetic amplifier will reflect a very high impedance into the gate circuit during mode II. If the magnetic amplifier has been connected in series with the load, as in one of the conventional circuits, the high impedance reflected from the control circuit into the gate windings during mode II will cause most of the applied a-c supply voltage to appear across the reactors.

Parallel Output

Fig. 2 shows the a-c self-saturating magnetic amplifier or "doubler." During mode II, with practically all of the a-c supply voltage appearing across the reactors, one reactor will be taken toward saturation, while the second reactor will be reset by the control voltage. During the next half-cycle, the volt-seconds absorbed from the a-c line by the reset reactor will be dependent upon the volt-seconds that had been reset into it by the control in the first half-cycle.

As noted previously, the output voltage across the series load element R_L is equal to the difference between the volt-seconds supplied by the line and the volt-seconds absorbed by the controlled element, the reactor. Also, it is interesting that, while variations in a-c voltage will affect the output, the average voltage per cycle dropped across the reactors is dependent only upon the control signal.[3] Since the average potential across the reactors is less subject to random variations than the potential across R_L, one is led to consider the possibilities of connecting the output load in parallel with the magnetic amplifier instead of in series; see Fig. 3(A).

The conclusions typified by Fig. 3 may also be reached by starting with the saturating voltage-stabilizing transformer; see Fig. 4(A). Saturating transformers using high-quality, square hysteresis loop core materials have been manufactured for some time. The transformer and a-c supply are deliberately proportioned so that the supply always furnishes more volt-seconds per half-cycle than the magnetic core of the transformer can support. The core of the saturating transformer therefore is driven in each half-cycle from negative to positive saturation (or vice versa) by the supply. The transformer core flux deviation per cycle becomes a fixed constant value. Consequently, the volt-seconds induced in the transformer secondary winding are independent of variations in load resistance or a-c supply voltage. Since the core of the transformer saturates before the end of each half-cycle, the saturating transformer appears as a dead short circuit twice every cycle. The element Z_L is added for the sole purpose of limiting these short-circuit currents.

The saturating transformer serves as a very useful and simple dynamically stable voltage source. However, its output is not readily controlled unless some method is found to adjust the total core flux

---

Paper 58-178, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 10, 1957; made available for printing November 26, 1957.

Allen B. Rosenstein is with the University of California, Los Angeles, Calif.

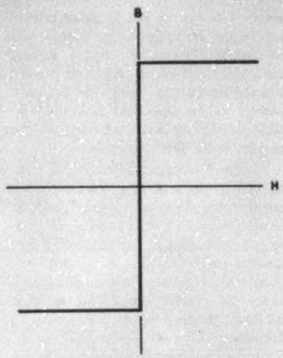


Fig. 1. Idealized B-H characteristics

change per half-cycle. To control the induced secondary volt-seconds per cycle, insert a diode in series with the primary of the saturating transformer. The a-c line can now cause a flux excursion in only one direction. To reset the core flux during the following half-cycle, add a control winding to the transformer, along with a control impedance and control voltage; see Fig. 4(B). The control winding will thus reset the core flux to any desired level. The control thereby governs or predetermines the volt-seconds, which may be coupled into the secondary during the following power half-cycle. To decouple the load from the flux-resetting control voltage, a diode is added in the transformer output winding. Again, the volt-seconds induced in the secondary will be independent of variations in the a-c supply voltage.

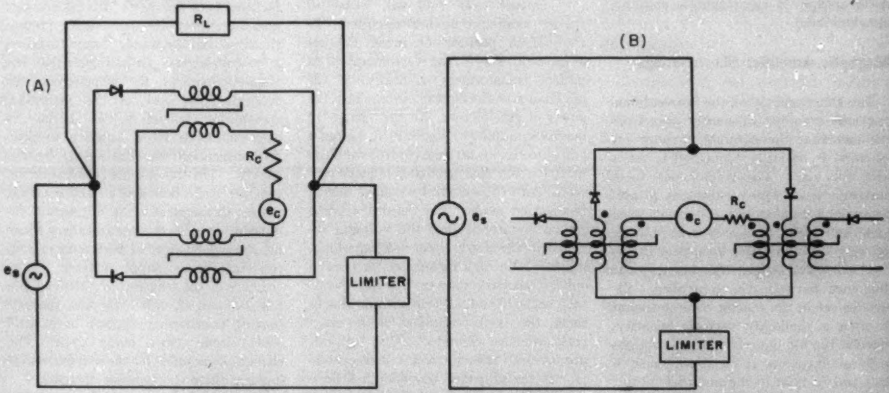


Fig. 3. A—Self-saturating a-c magnetic amplifier with shunt (parallel) loading. B—Full-wave transactor

The Transactor

Since the transformer now plays the dual role of a transformer and a saturable reactor, it is proposed that it be called a "transactor" and it shall be so referred to in the remainder of this paper. Z_L continues to play the role of a current limiter after the transactor has saturated; it shall hereafter be referred to as the "limiter."

The half-wave self-saturated transactor of Fig. 4(B) might be considered the fundamental building block of all transactor circuits. It will display a number of desirable characteristics, including:

1. Half-cycle response.
2. An output which will be largely independent of a-c supply variation.
3. Output circuit isolated from the a-c supply line (which means essentially magnetic coupling).
4. Output voltage independent of load impedance. (The transactor can work into an open circuit.)
5. Low quiescent output voltage which is independent of the a-c line and output loading.

In keeping with magnetic-amplifier tradition, the analysis of the half-wave transactor will be developed in Part II of this treatise. In Part I, a detailed analysis of a full-wave transactor system is presented.

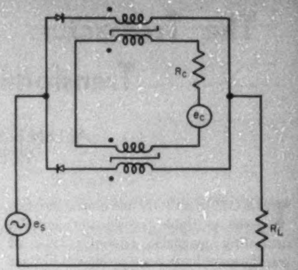


Fig. 2. Self-saturating a-c magnetic amplifier, doubler

The half-wave transactor of Fig. 4(B) is readily converted to a full-wave output by the addition of another saturable transformer and two rectifiers; see Fig. 3(B). Comparing the circuits of Figs. 3(A) and (B) it is seen that they are fundamentally the same. The transactor is intimately related to the shunt-loaded self-saturating a-c magnetic amplifier. In the transactor circuit, the load is transformer-coupled in parallel with the magnetic amplifier. Although the addition of the transformer secondary winding obviously reduces the power-handling capacity of a given core, it does offer certain advantages, including the convenience of interstage coupling and the

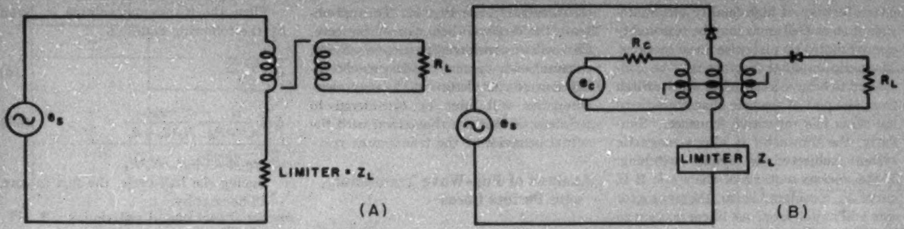


Fig. 4. A—Saturating transformer voltage regulator. B—Half-wave self-saturated transformer–half-wave transactor

elimination of the variations of IR (current-resistance) drop in the primary gate windings and rectifiers. The two transactor secondaries can also be readily connected in series, center tap, to give a full-wave d-c output, or in parallel to give a full-wave a-c output.

THE LIMITER

After the transactor saturates, some type of series-limiting device in series with the transactor primary is necessary to prevent the alternating current from reaching destructive values. Any of the linear circuit elements of resistance, capacitance, or inductance may be used. It is expected that for small-signal low-power transactors, resistors will be used as limiters for simplicity and minimum cost. Where higher frequency a-c carriers are employed, capacitors should be practical. In fact, with series capacitors, additional regulating action may be obtained by designing the system as a nonlinear series-resonating regulator. The transactor will resonate with the capacitor and saturate before the end of each half-cycle.

Where it is desirable to handle large amounts of power, the limitations of linear limiters are too well known to warrant further discussion. For the large power transactor the ideal limiter would be a nonlinear zero-power-dissipating limiter which would restrict the transactor primary current to instantaneous values only slightly in excess of those required for maximum power output. Several biased-reactor circuits suggest themselves, possessing characteristics closely approximating the ideal. If the limiters are properly constructed and biased, they will allow the transactor to operate from no load to full load without affecting the output voltage. However, the limiters would be set for a given maximum current, which would make the system self-protecting and even allow long-time short-circuiting of the transactor output.

Several variations in limiter circuitry will be described in Part II. For Part I the limiter will simply be an ordinary series-connected saturable reactor with very high control circuit impedance.

Circuit Analyzed

The full-wave transactor system which was analyzed and tested is given in Fig. 5. The behavior of the series-saturable reactor with high-control circuit impedance has been fully developed by others.[4] This series reactor has been shown to behave as a constant alternating current device. It supplies square waves of current of fixed magnitude to the load, which in this case is the transactor.

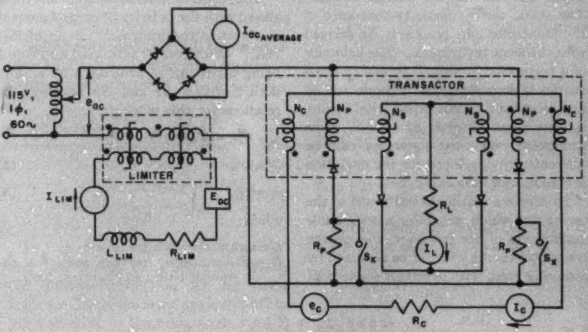


Fig. 5. Full-wave transactor schematic

The transactor shall act as a magnetically controlled saturable transformer and shall gate to the output a rectangular-pulse-width modulated wave. The width of the output wave shall be controllable from nearly zero to full half-cycle. For a given control setting, the output volt-seconds will remain constant under variation of load resistance a-c supply voltage, and limiter peak current. However, the height and consequently the width of the output pulse will vary although the area remains constant.

As a result of the choice of limiting circuits, for the purposes of this analysis, the alternating voltage source and the limiter may be replaced by an equivalent constant-current, square-wave generator. The choice of limiters will offer several advantages, not the least of these including:

1. An appreciable simplification of the complexity of the analysis.
2. A relatively linear transfer function when a d-c control signal is used.
3. A pulse-width modulated square-wave output which should be ideal for driving switching transistors or other stages of magnetic amplification.

Method of Analysis

It has been shown that the ultimate performance of the high-gain magnetic-amplification system is a direct function of its core material.[4] The static B-H characteristics of high-quality core materials such as Deltamax may be reasonably approximated by piecewise linear sections. Magnetic-amplifier reactors may be considered to behave as switches which switch from one part of the core characteristic to the other in a repetitive sequence. Similarly, the transactor is also a magnetic system subjected to cyclic switching. If the various sections of the static **B**-**H** curve are considered linear, the transactor core will switch from one linear portion to the other. The performance may then be represented by a sequence of linear equations.

Johannsen,[5] Bourne,[6] Hubbard,[7] Kikuchi,[8] Kusada[9] and others have successfully applied difference equations to magnetic-amplifier analysis. (Particular attention is called to Kusada's paper which analyzes the a-c doubler circuit.) The static and dynamic performance of the transactor can profitably be derived with the same techniques. The behavior of a transactor in any given half-cycle of output is dependent upon the amount of flux reset the core has experienced during the previous half-cycle of control. The transactor difference equations can be conveniently developed for the deviation in core flux for each half-cycle.

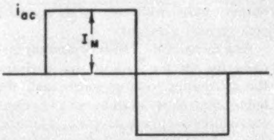


Fig. 6. Supply current wave form

To evolve a mathematical model of the transactor which will offer a reasonable explanation of its behavior, the chosen transactor circuit will first be analyzed by assuming cores with a perfect static **B**-**H** characteristic; see Fig. 1. To approximate the dynamic behavior of the core, they will be considered to possess a finite, constant eddy-current resistance. Modifications of the picture of the core characteristics will later be considered to reconcile further the theoretical with the actual behavior of the transactor.

Analysis of Full-Wave Transactor with Perfect Cores

Reference is made to the schematic. The limiter acts as a constant-current square-wave generator. The transactor will be analyzed on the basis of a constant-current square-wave source, see Fig. 6. The transactor core material is assumed to possess a perfectly square static **B**-**H** characteristic with zero hysteresis; see Fig. 1. The core material, however, possesses finite resistivity and therefore the dynamic **B**-**H** curve, because of core eddy currents, will have a finite separation between its ascending and descending legs.

Starting with zero bias current, both cores are always saturated if perfect **B**-**H** characteristics are assumed.

Mode I Operation

The control circuit switch may be closed with the polarity of control current as shown in the equivalent circuit of Fig. 7(A). With core II saturated and core I being taken out of saturation by e_c, and with rectifiers $S1$ and $S3$ blocked, the relations for this mode are:

$$e_c - N_c e_0 = i_c R_c \tag{1}$$

$$N_c i_c = i_e \tag{2}$$

$$e_0 = i_e R_e \tag{3}$$

where

e_0 = transactor volts per turn
R_e = effective eddy current resistance assuming 1-turn eddy current loop

Then, the transactor voltage is given by the following equation

$$e_0 = \frac{e_c/N_c}{A} \tag{4}$$

$$i_e = \frac{e_c}{AN_c^2R_c} = \frac{e_c}{N_c^2R_e + R_c} \tag{4A}$$

where $A = 1 + R_c/N_c^2R_e$.

During the half-cycle, the flux in core I will be reset by

$$\Delta_0 = \frac{1}{\omega}\int_0^{\pi} e_0 d\theta \tag{5}$$

Mode II Operation

When the polarity of the source current i_{ac} is reversed, the flux of core I will be driven toward saturation while the flux of core II will be taken out of saturation. With a small control voltage it will be assumed that rectifiers $S2$ and $S4$ must now block with rectifiers $S1$ and $S3$ conducting. The new equivalent circuit is shown in Fig. 7(B), the relations for which are

$$e_1 = i_{e_1} R_e \tag{6}$$

$$e_1 = \frac{i_{L_1} R_L}{N_s} \tag{7}$$

$$e_2 = i_{e_2} R_e \tag{8}$$

$$e_c + N_c e_1 - N_c e_2 = i_c R_c \tag{9}$$

$$N_c i_c - i_{e_2} = 0 \tag{10}$$

$$N_c i_c + i_{e_1} + N_s i_{L_1} - N_p i_1 = 0 \tag{11}$$

where e_1, e_2 = transactor volts per turn.

Since $i_1 = i_{ac}$, the six equations may be solved for e_1, e_2, i_c, i_{e_1}, i_{e_2}, i_{L_1}. Therefore during the interval in which the conditions of Fig. 7(B) prevail

$$e_2 = \frac{B}{1+AB}\,\frac{e_c}{N_c} + \frac{R_e N_p i_{ac}}{1+AB} \tag{12}$$

$$e_1 = -\frac{1}{1+AB}\,\frac{e_c}{N_c} + \frac{A}{1+AB}\,R_e N_p i_{ac} \tag{13}$$

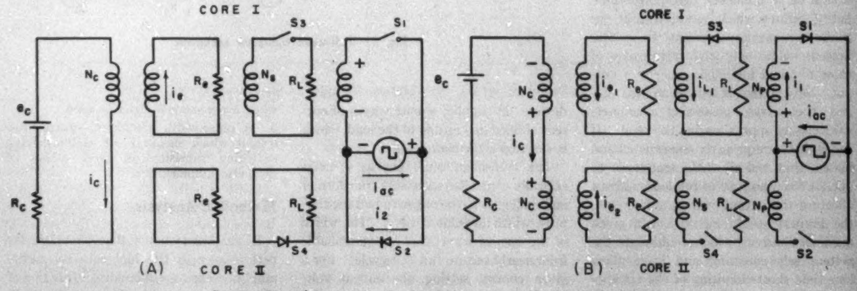


Fig. 7. Equivalent circuit. A—Mode I operation. B—Mode II operation

Vol.
Pt. 1
1958

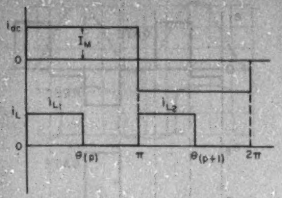


Fig. 8. Supply line I_{ac} and load I_L currents

$$i_{L_1}=-\frac{1}{1+AB}\frac{N_s}{N_cR_L}e_c+\frac{A}{1+AB}\frac{R_s}{R_L}N_sN_pi_{ac} \qquad (14)$$

$$i_c=\frac{B}{1+AB}\frac{e_c}{N_c^2R_s}+\frac{N_pi_{ac}}{N_c(1+AB)} \qquad (15)$$

$$i_{e_1}=-\frac{1}{1+AB}\frac{e_c}{N_cR_s}+\frac{A}{1+AB}N_pi_{ac} \qquad (16)$$

$$i_{e_2}=\frac{B}{1+AB}\frac{e_c}{N_cR_c}+\frac{N_pi_{ac}}{1+AB} \qquad (17)$$

where

$$A=1+\frac{R_c}{N_c^2R_s}$$

$$B=1+\frac{R_sN_s^2}{R_L}$$

When control voltage is applied, the flux level in each core is reset progressively lower each cycle until finally the minor loop reaches steady state.

It should be understood that in the course of each half-cycle, the mode of operation shown in Fig. 7(A) normally follows the mode illustrated in Fig. 7(B). At the beginning of each half-cycle, one core will be found saturated while the other will be unsaturated. Fig. 7(B) shows the originally saturated core (core II) being reset while the other (core I) is being carried by the load voltage toward saturation. Some time during the half-cycle, core I will saturate and short-circuit out all of its associated windings. The conditions of Fig. 7(A) will now prevail for the remainder of the half-cycle. For the Fig. 7(A) condition, however, it will be core I saturated and core II being reset. With the next half-cycle the role of the cores will be reversed: Δ_0 is obtained by integrating over a full half-cycle only for that initial half-cycle following application of e_c. For subsequent half-cycles the value of Δ_0 is found by integrating e_0 from $\theta(\rho)$ to π.

Although both equivalent circuits, Fig. 7(A) and (B), show R_L in two places, the actual circuit had only one R_L. The two R_L's are shown simply for convenience in drawing the equivalent diagrams. It is to be noted that only one transactor at a time will couple the load resistor R_L over to be fed from the current source. Fig. 8 shows current voltage waveforms for one cycle. Fig. 9 gives the dynamic **B-H** loop.

During one cycle the flux changes of one core are

$$\Delta_2=\frac{1}{\omega}\int_0^{\theta(p)}e_2d\theta \qquad (18)$$

$$\Delta_0=\frac{1}{\omega}\int_{\theta(p)}^{\pi}e_0d\theta \qquad (19)$$

$$\Delta_1=\frac{1}{\omega}\int_{\pi}^{\theta(p+1)}e_1d\theta \qquad (20)$$

Here the voltage e_2 is the reset potential across the core (corresponding to core II, Fig. 7(B)) during the first part of the cycle $0-\theta(p)$ before the other core has saturated; e_0 is the reset potential across the same core [corresponding to core I in Fig. 7(A)] during the rest of the half-cycle $\theta(p)-\pi$ after the other core has saturated. The setting potential is e_1 on the core [similar to core I, Fig. 7(B)] during the following half-cycle in the interval from π to $\theta(p+1)$ where the core saturates.

Since the total sum of the flux changes during one complete cycle is zero for both transient and steady-state operation,

$$\Delta_2+\Delta_0=\Delta_1 \qquad (21)$$

Substitute equations 4, 12, 13 into equations 19, 18, and 20; then into equation 21 (remembering that $e_c=E_c$ and $i_{ac}=I_M$).

$$\frac{B\dfrac{E_c}{N_c}+R_sN_pI_M}{1+AB}\int_0^{\theta(p)}d\theta+\frac{\dfrac{E_c}{N_c}}{A}\int_{\theta(p)}^{\pi}d\theta=\frac{-\dfrac{E_c}{N_c}+AR_sN_pI_M}{1+AB}\int_0^{\theta(p+1)}d\theta \qquad (22)$$

Letting $E_c/N_c=E_c'$ and $R_1N_pI_M=I_M'$, and performing the integration, yields the difference equation for the firing angle θ, relating the firing angle $\theta(p)$ of the pth half-cycle to the firing angle $\theta(p+1)$ for the $p+1$ half-cycle:

$$(A^2I_M'-AE_c')\theta(p+1)-(AI_M'-E_c')\theta(p)=(1+AB)\pi E_c' \qquad (23)$$

For the steady-state condition $\theta(p)=\theta(p+1)$

$$\theta(\infty)=\frac{(1+AB)\pi E_c'}{-(A-1)E_c'+A(A-1)I_M'}=\frac{\left(1+\dfrac{N_c^2R_s+R_c}{N_c^2R_s}+\dfrac{R_L+R_sN_s^2}{R_L}\right)\pi E_c'}{-\dfrac{R_c}{N_c^2R_s}E_c'+\dfrac{R_c}{N_c^2R_s}\dfrac{N_c^2R_s+R_c}{N_c^2R_s}I_M'} \qquad (24)$$

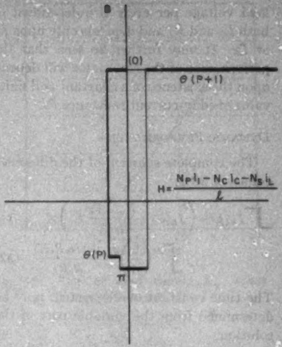


Fig. 9. Dynamic B-H loop

Under most operating conditions this reduces to

$$\theta(\infty)\cong k\frac{\pi E_c'}{I_M'} \qquad (25)$$

In any given half-cycle the average load current per cycle may be expressed as

$$I_L(p)=\left(\frac{N_s}{\pi R_L}\frac{AI_M'-E_c'}{1+AB}\int_0^{\theta(p)}d\theta\right)\text{(before saturation, equation 14)}+\left(0\int_{\theta(p)}^{\pi}d\theta\right)\text{(after saturation)} \qquad (26)$$

$$I_L(p)=\frac{N_s}{\pi R_L}\frac{AI_M'-E_c'}{1+AB}\theta(p) \qquad (27)$$

$$I_L(p+1)=\frac{N_s}{\pi R_L}\frac{AI_M'-E_c'}{1+AB}\theta(p+1) \qquad (28)$$

Substituting the values for $\theta(p)$ and $\theta(p+1)$ into equation 23 yields

$$\frac{R_L}{N_s}AI_L(p+1)-\frac{R_L}{N_s}\theta(p)=E_c'$$

$$I_L(p+1)-\frac{1}{A}I_L(p)=\frac{N_sE_c}{N_cR_LA} \qquad (29)$$

The difference equation for the output current.

STEADY-STATE PERFORMANCE

Under steady-state conditions $I(p+1)=I_{(p)}$

$$I_L(\infty)=\frac{N_cN_sR_s}{R_cR_L}E_c \qquad (30)$$

$$=\frac{N_cN_sR_s}{R_L}I_c$$

$$E_L(\infty)=N_cN_sR_sI_c \text{ as } E_L=R_LI_L \qquad (31)$$

It is seen that the steady-state average load voltage per cycle is independent of both I_M and R_L and depends only upon E_c or I_c. It may further be seen that the performance of the transactor will depend upon the existence of a constant and finite value of eddy-current resistance R_e.

DYNAMIC PERFORMANCE

The complete solution of the difference equation 29 is

$$\lrcorner\!\ulcorner I_L(p)=\left(I_L(0)-\frac{N_cN_sR_e}{R_cR_L}E_c\right)\times \lrcorner\!\ulcorner\left(\frac{1}{A}\right)^p+\frac{N_cN_sR_eE_c}{R_cR_L} \qquad (32)$$

The time constant of the system may be determined from the variable part of the solution.

$\left(\frac{1}{A}\right)=e^{-\alpha\frac{1}{2f}}$ since the duration of each interval is $\frac{1}{2f}$ and $t=\frac{p}{2f}$

The system time constant

$$\tau=\frac{1}{\alpha}=\frac{1}{2f\log(A)}$$

$$=\frac{1}{2f\log\dfrac{N_c^2R_e+R_c}{N_c^2R_e}}\cong\frac{N_c^2R_e}{2fR_c} \qquad (33)$$

It is seen that the time constant of the system may be reduced by decreasing the eddy-current resistance R_e. However, from equation 30, it may be further seen that the current gain also varied directly with R_e.

If power gain k_p is defined as

$$k_p=\frac{I_L^2R_L}{I_c^2R_c}k_f^2$$

where k_f is the form factor of i_L. Then

$$k_p=\frac{N_c^2N_s^2R_e^2}{R_LR_c} \qquad (34)$$

From equations 30, 33, and 34 the dynamic power gain D_p becomes, if

$$D_p=\frac{k_p}{\tau}$$

$$D_p=2f\frac{N_s^2}{R_L}\times R_e\times k_f^2 \qquad (35)$$

If D_p can be considered as a figure of merit, it becomes apparent that the performance of the transactor will be materially improved by using a high-frequency carrier and a high-resistivity core material such as a ferrite.

CORRELATION OF EXPERIMENTAL FACT WITH THEORETICAL FANCY

The theoretical waveshapes of the important voltages, currents and fluxes are drawn in Fig. 10. Starting at the top of the page with the sinusoidal line voltage, it is followed by the square wave of the a-c line current generated by the limiter. The phase angle of i_{ac} lags e_{ac}. However, as the firing angle $\theta(p)$ of the load current i_L is shifted back toward π, the phase difference between e_{ac} and i_{ac} is reduced. This is the result of the load's absorbing more energy and causing the constant current source effectively to supply a larger resistance. The voltage across the limiter is the difference between e_{ac} and N_p/N_se_L. The behavior of the control current i_{c1} is governed by equations 4(A) and 15. The gating portions of e_1 and e_2 are shown positive while the reset voltages are negative. Their magnitudes are given by equations 4, 12, and 13. The core fluxes Φ_1, and Φ_2 are the integrals of e_1 and e_2.

Fig. 11 gives a series of oscillographs of the voltages and currents whose theoretical behavior is depicted in Fig. 10. The true waveforms offer reasonably close agreement with the derived waveshapes.

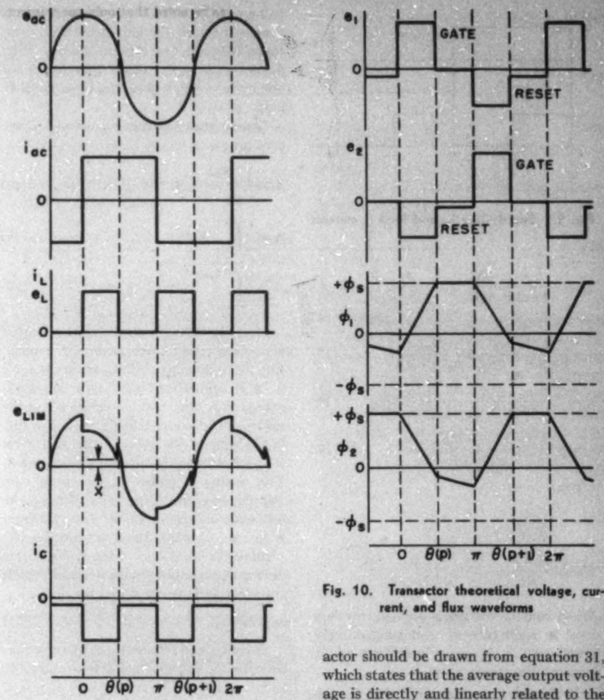


Fig. 10. Transactor theoretical voltage, current, and flux waveforms

The conclusions regarding the steady-state theoretical performance of the transactor should be drawn from equation 31, which states that the average output voltage is directly and linearly related to the average control voltage. It is further implied that within its design range this linear relation between E_L and E_c is completely independent of variations in e_{ac}, I_{lim}, R_L, and supply frequency.

A glance at Fig. 12 immediately shows that the projection of I_L versus I_c does not go to zero when $I_c=0$. Instead, the projection of I_L versus I_c gives a substantial finite value of I_c for $I_L=0$. Since it is known that the static hysteresis loop does in reality have a finite width, this result is not too surprising. It will be treated in a later section when the static **B-H** loop is modified for a finite separation of the ascending and descending legs.

A judicious or perhaps a fortunate choice of simplifying assumptions will often allow the principal experimental results to offer a reasonable resemblance to those predicted theoretically. The choice of **B-H** characteristics has been fortunate in that the test data show excellent correlation of the first-order effects. Fig. 12 has the predicted linear relation between control and load current with the noted offset. Table I lists the output current for three values of con-

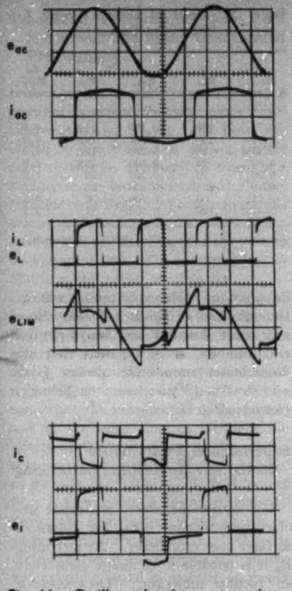


Fig. 11. Oscillographs of transactor voltage and current waveforms

Table I. Effect of Line Voltage e_{ac} Upon Load Current I_L

Load Resistance and I_{lim} Held Constant

e_{ac}, Volts	I_L, Amperes	I_c, Amperes
90	0.550	0.008
70	0.550	
50	0.550	
40	0.550	
32	0.540	
90	0.212	0.006
70	0.213	
50	0.214	
40	0.215	
21	0.217	
16.5	0.228	
90	0.097	0.004
70	0.098	
50	0.098	
40	0.098	
30	0.097	
20	0.09	

R_L = 50 ohms; I_{lim} = 0.30 ampere; R_c = 10,000 ohms.

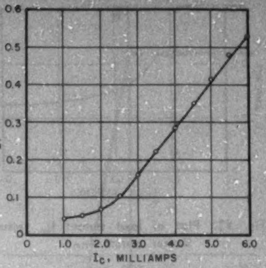


Fig. 12. Plot of load current I_L, versus control current, I_c

R_L = 50 ohms
R_s = 400 ohms
R_p → 0 ohms

trol current with the load resistance and I_{lim} held constant. For each value of I_c, the a-c line voltage was varied 300% to 500%. This covered a range from its maximum allowable value of 90 volts down to a minimum which gave zero residual voltage drop across the limiter. Table II offers the same type of information except that the line voltage is held constant and the effect of a 500% variation of the limiter current is presented. Table III shows the effects upon the output voltage of a load resistance variation of from 25 ohms to infinity. The readings here have been taken for initial open-circuit outputs of 15 and 20 volts. The effects of different values of R_c are also given.

Although the first-order experimental performance agrees quite well with that predicted by theory, closer examination reveals the usual second-order flies in our theoretical ointment. In Table III for R_L under 250 ohms when R_c is greater than 10,000 ohms, the output voltage is found actually to increase with decreasing load resistance. Even Table II with $R_c = 10K$ shows an increasing output current over part of the range of decreasing I_{lim}.

Since the deviation from predicted performance appeared to be a function of R_c (and consequently of E_c) another check was made of I_L versus I_{lim} with $R_c = 13K$, see Fig. 13. Here it was found that the load current would "run away" with decreasing I_{lim}. Further consideration of Fig. 7(A) gave the clue to the reason for this behavior when it became clear that with saturation of the gating core, a large value of e_c would unblock the nonconducting primary rectifier, in this case S_1 during mode I operation. To explain this unexpected variation of I_L with I_{lim}, it is necessary to consider mode I primary rectifier unblocking and consequently the resistance of the primary winding R_p. Referring to the circuit schematic, Fig. 5, switches S_x were opened, inserting the resistance R_p into the transactor primary circuit.

Transactor Performance with Rectifier Unblocking

The original assumptions included the statement that e_c is considered small, and therefore that rectifier S_1 is blocked during the saturated interval of core II. Further study indicates that this statement is true only if $R_p \neq 0$ and if the voltage $I_{ac}R_p$ (N_c/N_p) is greater than $N_c e_0$ or from equation 4.

$$\frac{N_c i_{ac} R_p}{N_p} \geq N_c e_0 = e_c/A = \frac{e_c N_c^2 R_s}{N_c^2 R_s + R_c} \qquad (36)$$

Since, for a given output e_c varies with R_s and R_c, the conditions of equation 36 can be readily upset by increasing R_c or decreasing R_s (common and often desirable variations in the system parameters). The system's behavior must therefore be re-examined during the saturation interval, i.e., second part of the reset interval.

Table II. Effect of Limiter Current I_{lim} Upon Load Current I_L

I_{lim}, Amperes	I_L, Amperes	I_c, Amperes
0.10	0.234	0.006
0.15	0.229	
0.17	0.222	
0.20	0.217	
0.30	0.217	
0.40	0.228	
0.50	0.246	

Note: e_{ac} = 70 volts; R_L = 50 ohms; R_c = 10,000 ohms.

Table III. Effect of Load Resistance R_L Upon Output Voltage E_L for Various Values of Control Resistance R_c

Ohms		E_L, Volts						
R_L	∞	15.0	15.0	15.0	15.0	20.0	20.0	20.0
	500	13.7	14.0	14.0	13.7	18.8	18.8	18.5
	250	13.5	13.7	13.7	13.5	18.6	18.6	18.4
	100	12.5	12.3	11.7	10.2	19.0	18.6	18.0
	50	12.7	12.0	10.5	7.8	21.5	21.0	19.5
	25	14.7	12.8	10.2	6.0	20.5	20.5	20.5
R_c		15K	12.5K	10K	5K	15K	13.5K	12.5K

e_{ac} = 70 volts; I_{lim} = 0.30 ampere.

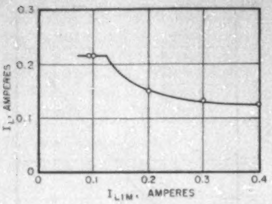


Fig. 13. Plot of load current I_L, versus limiter current I_{lim}

$R_L = 100$ ohms
$R_e = 13$K ohms
$R_p \to 0$ ohms
$E_{AC} = 70$ volts

Assuming that the conditions of equation 36 are not satisfied, Fig. 7(A) must be redrawn; see Fig. 14, with rectifier S_1 unblocked.

Writing the system equations yields

$$e_c - N_c e_0 = i_c R_c \qquad (37)$$

$$N_c i_c - i_e - N_p i_1 = 0 \qquad (38)$$

$$e_0 = i_e R_e \qquad (39)$$

$$e_0 = \frac{i_{ac} R_p + 2 i_1 R_p}{N_p} = \frac{(i_{ac} + 2 i_1) R_p}{N_p} \qquad (40)$$

Solving the four simultaneous equations for the reset voltage e_0 and the control current i_c during the interval yields

$$e_0 = D N_p i_{ac} + 2D \frac{N_c}{R_c} e_c \qquad (41)$$

where

$$D = \frac{1}{2\frac{N_c^2}{R_c} + \frac{2}{R_e} + \frac{N_p^2}{R_p}}$$

$$i_c = \frac{1 - 2DN_c^2}{R_c} e_c - D N_c N_p i_{ac} \qquad (42)$$

Note: $D \to 0$ as $R_p \to 0$

It may be noted in passing that i_c cannot reverse direction and become negative since rectifier S_1 would block, and conditional equation 36 would apply.

Modifying equation 21 with the new value of e_0 (equation 41) yields

$$\frac{B\frac{E_c}{N_c} + R_c N_p I_M}{1+AB} \int_0^{\theta(p)} d\theta + \left(\frac{D}{R_e} N_p R_e I_M + 2D \frac{N_c^2}{R_c} \frac{E_c}{N_c}\right) \int_{\theta(p)}^{\pi} d\theta = \frac{-\frac{E_c}{N_c} + A R_e N_p I_M}{1+AB} \int_0^{\theta(p+1)} d\theta \qquad (43)$$

Letting $E_c/N_c = E_c'$ and $R_e N_p I_M = I_M'$ and integrating, yields the following

$$\frac{BE_c' + I_M'}{1+AB}\theta(p) + \left(\frac{D}{R_e} I_M' + 2D \frac{N_c^2}{R_e} E_c'\right) \times (\pi - \theta(p)) = \frac{AI_M' - E_c'}{1+AB} \theta(p+1)$$

$$(AI_M' - E_c')\theta(p+1) - \left[BE_c' + I_M' - (1+AB)\left(\frac{D}{R_e} I_M' + 2D \frac{N_c^2}{R_c} E_c'\right)\right] \times \theta(p) = (1+AB)\pi\left(\frac{D}{R_e} I_M' + 2D \frac{N_c^2}{R_e} E_c'\right) \qquad (44)$$

Substituting equations 27 and 28 for $\theta(P)$ and $\theta(p+1)$ into equation 44 yields the difference equation for the output load current

$$I_L(p+1) - \frac{BE_c' + I_M' - (1+AB) \times \left(\frac{D}{R_e} I_M' + 2D \frac{N_c^2}{R_c} E_c'\right)}{AI_M' - E_c'} I_L(p) = \frac{N_s}{R_L}\left(\frac{D}{R_e} I_M' + 2D \frac{N_c^2}{R_c} E_c'\right) \qquad (45)$$

Steady state occurs when $I_L(p+1) = I_L(P)$

$$I_L(\infty) = \frac{\frac{DN_s}{R_L}\left(\frac{I_M'}{R_e} + \frac{2N_c^2}{R_c} E_c'\right)(AI_M' - E_c')}{\left(A - 1 + (1+AB)\frac{D}{R_e}\right) I_M' - \left(1 + B - (1+AB) 2D \frac{N_c^2}{R_c}\right) E_c'} \qquad (46)$$

For small values of R_p, and consequently D, equation 46 becomes

$$I_L(\infty) = \frac{\frac{DN_s}{R_L}(AI_M' - E_c')}{(A-1)I_M' - (1+B)E_c'} \times \left(\frac{I_M'}{R_e} + \frac{2N_c^2}{R_c} E_c'\right) \qquad (47)$$

Table IV. Effect of Limiter Current I_{lim} Upon Load Current I_L

I_L, Amperes	R_L, Ohms	I_{lim}, Amperes	R_e, Ohms	I_c, Milliamperes
0.1	100	0.200	13K	4.9
0.16	100	0.500	13K	4.9
0.24	25	0.200	13K	4.9
0.38	25	0.500	13K	4.9
0.20	25	0.200	20	1.5
0.4	25	0.500	20	1.5
0.11	100	0.200	20	2.3
0.22	100	0.500	20	2.3

Resistance $R_p = 10$ ohms (no rectifier unblocking).

Obviously, both the numerator and denominator difference terms will go through zero as I_M' is reduced. However, since A and B are always positive real numbers, it is apparent that the denominator term must always go to zero first and thus cause the behavior plotted in Fig. 13.

Experimental Performance Revisited: No Rectifier Unblocking

The deleterious effects of rectifier unblocking attendant upon low values of R_p have been established. By increasing R_p it is possible to eliminate completely any rectifier unblocking. The transactor should then perform as originally predicted by equation 31. Table IV gives the effect of the variation of I_{lim} upon I_L with $R_p = 10$ ohms for several values of R_L and R_e. It is readily observed that while the output no longer increases with decreased I_{lim}, I_L is unfortunately not independent of the limiter current. Some testing of the system with $x=0$, Fig. 10, (peak value of I_M may be held constant after saturation, but varied with E_{ac} during the gating period) indicates that the dependence upon I_{lim} occurs entirely during the gating interval. After the transactor saturates, the magnitude of I_{lim} has no further effect.

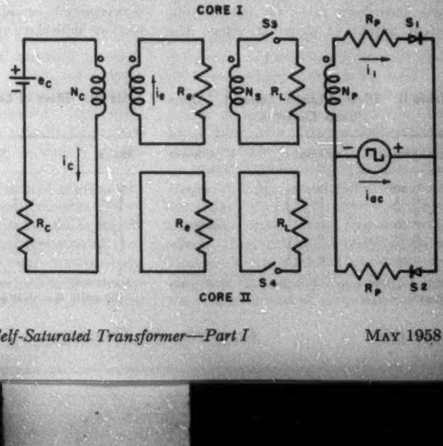


Fig. 14. Equivalent circuit considering R_p with core II saturated

Transactor Performance with Core Hysteresis and Finite Unsaturated Permeability

To reconcile the difference between the behavior experienced with no mode I rectifier unblocking, $R_p \neq 0$, and that predicted by the transactor with a perfect core, it will be necessary to consider a slightly more sophisticated **B-H** characteristic. In Fig. 15(A), static hysteresis exists, and the incremental permeability of the core in the unsaturated state is no longer infinite. The saturated permeability is still taken equal to zero.

Introduction of a finite, constant incremental permeability into the unsaturated conditions will cause the equivalent circuits to show N_p of both core I and core II shunted by a linear magnetizing inductance L_p. To reduce somewhat the complexity of the system and at the same time to highlight the effects of the finite permeability of the core, assume a static **B-H** characteristic, as given in Fig. 15(B). The **B-H** loop of Fig. 7(A) exhibits a finite permeability during the reset half-cycle and an infinite permeability during the setting half-cycle. Although the **B-H** characteristic chosen will never be found in practice, it should serve for the model because during the setting half-cycle, the magnetizing inductance of the gating core will appear across the current source and will not affect the number of volt seconds transformed by the core to the load. For further simplification, R_c shall be assumed to be infinite.

Based upon the **B-H** curve of Fig. 15(B), the setting core will appear as a perfect transformer while the resetting core will have a finite inductance L_p, in parallel with N_p, see Fig. 16. The rectifiers of the resetting core are assumed blocked.

Writing the system equations and taking the Laplace transform, yields

$$e_1(s) - \frac{i_{L_1}(s)R_L}{N_s} = 0 \tag{48}$$

$$e_2(s) - \frac{i_{p2}(s)L_p s}{N_p} + \frac{i_{p2}(0)L_p}{N_p} = 0 \tag{49}$$

$$N_c e_1(s) - N_c e_2(s) + e_c(s) - i_c(s)R_c = 0 \tag{50}$$

$$N_s i_{L_1}(s) - N_p i_{ac}(s) + N_c i_L(s) = 0 \tag{51}$$

$$-N_p i_{p2}(s) + N_c i_c(s) = 0 \tag{52}$$

Considering only the third term of equation 49, leads to the conclusion that

$$\frac{i_{p2}(0)L_P}{N_p} = 0 \tag{53}$$

because, during the latter part of the previous half-cycle when core II was saturated, L_{p2} vanished and therefore $i_{p2}(0) = 0$ at the beginning of the gating cycle.

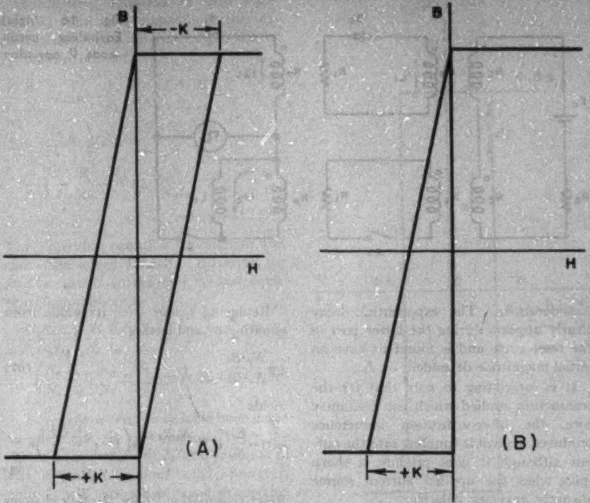


Fig. 15. B-H characteristics with hysteresis and finite unsaturated permeability

Solving the set of simultaneous equations yields

$$\left[\frac{N_s^2}{R_L} + \left(\frac{N_s^2}{R_L}\,\frac{R_c}{N_c^2} + 1\right)\frac{N_p^2}{L_p s}\right] e_2(s) = N_p I_{ac}(s) + \frac{N_s^2}{R_L N_c}\, e_c(s) \tag{54}$$

$$e_1(s) = -\frac{e_c}{N_c} + \left[1 + \frac{R_c}{N_c^2}\,\frac{N_p^2}{L_p s}\right] e_2(s) \tag{55}$$

$$i_{L_1}(s) = \frac{N_s e_1(s)}{R_L} \tag{56}$$

$$i_{p2}(s) = \frac{N_p}{L_p s}\, e_2(s) \tag{57}$$

Let

$$a = \frac{N_s^2}{R_L}$$

$$b = \left(\frac{N_s^2}{R_L}\,\frac{R_c}{N_c^2} + 1\right)\frac{N_p^2}{L_p}$$

$$\alpha = \frac{N_p I_M}{a} + \frac{E_c}{N_c} \tag{58}$$

and take the inverse transform of equations 54, 55, 56, and 57, remembering that

$$e_c(s) = \frac{E_c}{s} \text{ and } I_{ac}(s) = \frac{I_M}{s}$$

Then

$$e_2(t) = \alpha \epsilon^{-\frac{b}{a}t} \tag{59}$$

$$e_1(t) = -\frac{E_c}{N_c} + \frac{\dfrac{R_L}{N_s^2}}{R_c/N_s^2 + R_L/N_s^2}\,\alpha\epsilon^{-\frac{b}{a}t} + \frac{\dfrac{R_c}{N_c^2}}{R_c/N_c^2 + R_L/N_s^2}\,\alpha \tag{60}$$

$$i_{L_1}(t) = \frac{N_s}{R_L}\left(-\frac{E_c}{N_c} + \frac{R_L/N_s^2}{\dfrac{R_c}{N_s^2} + \dfrac{R_L}{N_s^2}}\,\alpha\epsilon^{-\frac{b}{a}t} + \frac{R_c/N_c^2}{\dfrac{R_c}{N_c^2} + \dfrac{R_L}{N_s^2}}\,\alpha\right) \tag{61}$$

$$i_{p2}(t) = \frac{a\alpha}{b}\left(1 - \epsilon^{-\frac{b}{a}t}\right)$$

$$i_{p2}(\theta_p) = \frac{a\alpha}{b}\left(1 - \epsilon^{-\frac{b}{a}\,\frac{\theta(p)}{\omega}}\right) \tag{62}$$

When $t = \theta(p)/\omega$ core I saturates and only E_c is available to continue to reset core II during the remainder of the half cycle; see Fig. 17(A) and (B).

From the reduced equivalent circuit, Fig. 17(B), it may be seen that for the second portion of the reset cycle, the voltage e_0 across core II will exponentially decay toward zero as i_c approaches its terminal value E_c/R_c. This conclusion is borne out experimentally when the voltage across winding N_c of the transactors is examined; see Fig. 18, which further justifies the choice of **B-H** characteristics and the inclusion of L_p in the system

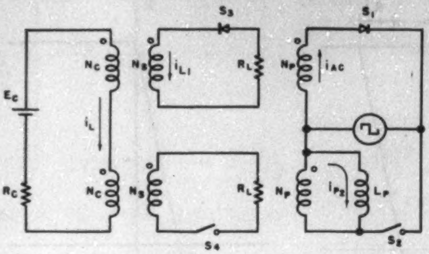


Fig. 16 (right). Equivalent circuit, mode II operation

considerations. The exponential decay clearly appears during the latter part of the reset cycle and is found to have an initial magnitude dependent on I_{lim}.

It is interesting to note that for the transactors studied which use Deltamax core, the after-saturation inductance produced negligible coupling into the output although it does produce a sharp spike when the applied current source abruptly reverses polarity.

Writing the transformed loop voltage and current equations for Fig. 17(A) with $t=0+$ at the instant-when the firing angle equals $\theta(p)$, yields

$$e_c(s)-N_c e_0(s)-i_c(s)R_c=0 \tag{63}$$

$$N_c i_c(s)-N_p i_{p_2}(s)=0 \tag{64}$$

$$e_0(s)-\frac{L_p s i_{p_2}(s)}{N_p}+\frac{L_p i_{p_1}\theta(p)}{N_p}=0 \tag{65}$$

Solving these simultaneous equations yields

$$e_0(s)=\frac{N_c L_p s}{N_c^2 L_p s+N_p^2 R_c}\,e_s(s)-\frac{N_p R_c L_p}{N_c L_p s+N_p^2 R_e}\,i_{p_1}(\theta_p) \tag{66}$$

Taking the inverse Laplace yields

$$e_0(t)=E_c/N_c\epsilon^{-\frac{N_p^2R_c}{N_c^2L_p}t}-\frac{N_pR_c}{N_c^2}i_{p_2}(\theta_p)\epsilon^{-\frac{N_p^2R_c}{N_c^2L_p}t}$$

Replacing $i_{p_1}(\theta p)$ with its value from equation 62 and letting

$$g=\frac{N_p^2R_c}{N_c^2L_p} \tag{67}$$

yields

$$e_0(t)=\frac{E_c}{N_c}\,\epsilon^{-gt}\,\frac{N_pR_c}{N_c^2}\,\frac{a}{b}\,\alpha\left(1-\epsilon^{-\frac{b}{a}\,\frac{\theta(p)}{\omega}}\right)\epsilon^{-gt} \tag{68}$$

Refer to equation 21 for the flux changes during one cycle:

$$\Delta_2+\Delta_0=\Delta_1 \tag{21}$$

Then

$$\frac{1}{\pi/\omega}\int_0^{\frac{\theta(p)}{\omega}}e_2dt+\frac{1}{\pi/\omega}\int_{\theta(p)/\omega}^{\pi/\omega}e_0dt=\frac{1}{\pi/\omega}\int_0^{\frac{\theta(p+1)}{\omega}}e_1dt \tag{21A}$$

Since, from equation 56, the output current $i_{L_1}=N_se_1/R_L$, the average value $I_{L_1}(p)$ of the output current during the pth half-cycle, may be obtained by integrating i_{L_1} over the half-cycle and dividing by π/ω.

$$I_L(p)=\frac{\omega}{\pi}\,\frac{N_s}{R_L}\int_0^{\frac{\theta(p)}{\omega}}e_1(t)dt \tag{69}$$

Multiplying both sides of equation 21(A) by N_s/R_L would cause the right side to become $I_L(p)$. However, the labor involved will be slightly reduced if, before making the substitution, $e_2(t)$ is rewritten in terms of $e_1(t)$. From equation 59 take $\alpha\epsilon^{-b/at}=e_2(t)$ and substitute directly for the like quantity in equation 60 to obtain

$$e_1(t)=\frac{\frac{R_L}{N_s^2}}{\frac{R_c}{N_c^2}+\frac{R_L}{N_s^2}}\,e_2(t)+\frac{\frac{R_c}{N_c^2}}{\frac{R_c}{N_c^2}+\frac{R_L}{N_s^2}}\,\alpha-\frac{E_c}{N_c}$$

Then

$$e_2(t)=\frac{\frac{R_c}{N_c^2}+\frac{R_L}{N_s^2}}{\frac{R_L}{N_s^2}}\,e_1(t)-\frac{R_c}{N_c^2}\,a\alpha+\frac{\frac{R_c}{N_c^2}+\frac{R_L}{N_s^2}}{\frac{R_L}{N_s^2}}\,\frac{E_c}{N_c} \tag{70}$$

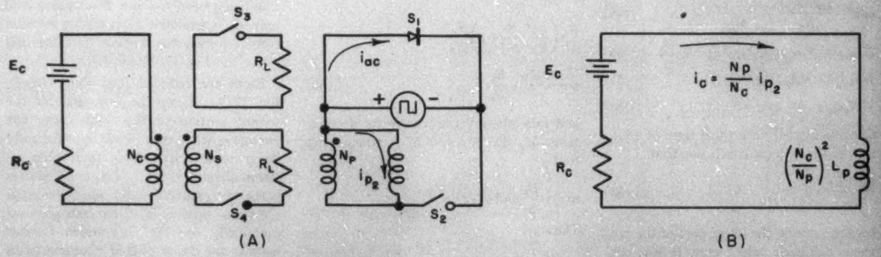
The next step is to multiply equation 21(A) by N_s/R_L and substitute in equations 70, 68, and 69, which will result in a difference equation in the dependent variable $I_L(p)$ and $\theta(p)$.

$$\frac{bL_p}{N_p^2}I_L(p)+\frac{N_s}{\pi R_L}\left(\frac{E_c}{N_c}-\frac{R_c}{N_c^2}N_pI_M\right)\theta(p)-\frac{\omega}{\pi}\,\frac{N_s}{R_{L\theta}}\left[\frac{E_c}{N_c}-\frac{N_pR_c}{N_c^2}\,\frac{a}{b}\,\alpha\left(1-\epsilon^{-\frac{b}{a}\,\frac{\theta(p)}{\omega}}\right)\right]\times\left(e^{-g\frac{\pi}{\omega}}-\epsilon^{-g\frac{\theta(p)}{\omega}}\right)=I_L(p+1) \tag{71}$$

To obtain a second equation in $I_L(p)$ and $\theta(p)$ the integration of equation 69 is performed after replacing $e_1(t)$ with its value from equation 60:

$$I_L(p)=\frac{\omega}{\pi}\,\frac{N_s}{R_L}\left(-\frac{E_c}{N_c}+\frac{R_c/N_c^2}{\frac{R_c}{N_c^2}+R_L/N_s^2}\,\alpha\right)\times\frac{\theta(p)}{\omega}-\frac{\omega}{\pi}\,\frac{N_s}{R_L}\,\frac{a}{b}\,\alpha\left(\frac{R_L/N_s^2}{R_c/N_c^2+R_L/N_s^2}\right)\times\left(\epsilon^{-\frac{b}{a}\,\frac{\theta(p)}{\omega}}-1\right) \tag{72}$$

Equations 71 and 72, containing both



(A)

(B)

Fig. 17. A—Equivalent circuit, mode I operation. B—Reduced equivalent circuit

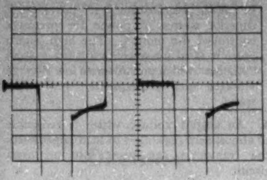


Fig. 18. Oscillograph of transactor reset voltage-expanded scale

first-order and exponential terms involving $\theta(p)$, are not readily solved directly. However, if the exponentials can justifiably be replaced by the first two terms of their expansions, the equations are readily reduced. The smaller the exponents of the exponential become, the more nearly the first two terms of the expansion approach the original function. The exponents under consideration are:

$$\frac{b}{a}\frac{\theta(p)}{\omega}=\frac{N_p^2}{L_p}\left(\frac{R_c}{N_c^2}+\frac{R_L}{N_s^2}\right)\frac{\theta(p)}{\omega}$$

and

$$g\frac{\theta(p)}{\omega}=\frac{N_p^2}{L_p}\frac{R_c}{N_c^2}\frac{\theta(p)}{\omega}$$

Using Arnold Engineering Company's published data for the slope of the Deltamax static **B**-**H** curve, the L_p of the experimental transactor is computed to be 25 henrys. For a $\theta(p)=\pi/2$, an $R_c \cong 1{,}800$ ohms will give an exponent in the vicinity of 0.1. For this range of exponents the exact value of the exponential differs from the value computed from the first two terms of its expansion by less than 0.5%.

Expanding the exponential of equation 72

$$\epsilon^{-\frac{b}{a}\frac{\theta(p)}{\omega}}=1-\frac{b}{a}\frac{\theta(p)}{\omega} \tag{73}$$

Substituting equation 73 into 72 yields

$$I_L(p)\cong\frac{N_s}{\pi R_L}\left(-\frac{E_c}{N_c}+\frac{R_c/N_c^2}{R_c/N_c^2+R_L/N_s^2}\alpha\right)\times \theta(p)+\frac{N_s}{\pi R_L}\frac{R_L/N_s^2}{R_c/N_c^2+R_L/N_s^2}\alpha\theta(p)$$

which reduces to

$$I_L(p)\cong\frac{N_p}{\pi N_s}I_M\theta(p)$$

and (74)

$$\theta(p)=\frac{\pi N_s}{N_pI_M}I_L(p)$$

Substituting equation 74 into equation 71 and also expanding the exponentials gives a difference equation for I_L, the load current.

$$\frac{N_s}{R_L}\frac{E_c}{N_c}+I_L(p)-\frac{N_s}{R_L}\frac{R_c}{N_c^2}\frac{\pi}{\omega}\left(\frac{N_s}{N_c}\frac{E_c}{I_M}+\frac{N_p}{N_s}R_L\right)I_L(p)+\frac{N_s^2}{R_L}\frac{R_c}{N_c^2}\frac{\pi}{\omega}\times\left(\frac{N_s}{N_c}\frac{E_c}{I_M}+\frac{N_p}{N_c}R_L\right)\frac{I_L^2(p)}{N_pI_M}=I_L(p+1) \tag{75}$$

This difference equation is nonlinear and cannot be solved directly. However, for steady state, $I_L(p)=I_L(p+1)$, and equation 75 simplifies to

$$\frac{N_sR_c}{2fN_cN_pI_M}\left(\frac{N_s}{N_c}\frac{E_c}{I_M}+\frac{N_p}{N_s}R_L\right)I_L^2(p)-\frac{R_c}{2fN_c}\times\left(\frac{N_s}{N_c}\frac{E_c}{I_M}+\frac{N_p}{N_s}R_L\right)I_L(p)+E_c=0 \tag{76}$$

Too many assumptions have been made to allow the use of equation 76 for the direct computation of I_L as a function of I_M. Equation 76 does prove, however, that I_L will increase in some proportion (but not quite in direct proportion) with increasing I_M.

When R_p is very small and R_c is large, the primary rectifier will unblock and the output will decrease as I_{1lm} is increased. On the other hand, with no rectifier unblocking, the opposite effect is true. By adjusting R_p and R_c it is possible to limit the output variation to a few per cent even with a 100% variation of I_{1lm}. It should be further noted that in practice it is usually a relatively simple matter to regulate I_{1lm} closely.

Transactor Performance with Core Hysteresis and Infinite Unsaturated Permeability

Equation 31, which was derived assuming a perfect **B**-**H** characteristic, predicts that the output load current I_L will be zero when $E_c=0$ and will increase linearly with E_c. Examining a typical variation of I_L versus I_c, Fig. 12, it is seen that the output remains constant at a low value until E_c reaches a rather substantial figure, and only then does I_L vary with I_c. The answer to this discrepancy lies again in the **B**-**H** loop assumptions. If it is assumed now that the core material possesses a static hysteresis characteristic with vertical sides, Fig. 19, the experimental behavior can be readily explained. K, the static coercive magnetomotive forces in ampere turns equals H_cl_1.

The equivalent circuit before saturation will be as shown in Fig. 7(B). The circuit equations will be the same as previously written except for the factor K in the sum of the ampere-turn equations.

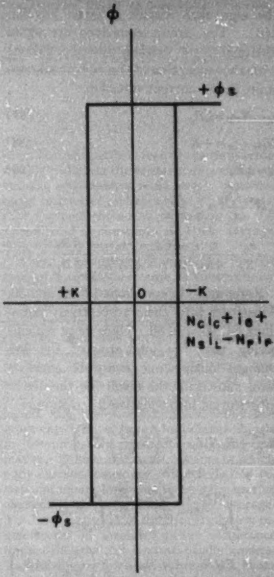


Fig. 19. B-H characteristic with hysteresis and infinite unsaturated permeability

$$e_1=i_{e1}R_e \tag{77}$$

$$e_1=\frac{i_{L1}R_L}{N_s} \tag{78}$$

$$e_2=i_{e2}R_e \tag{79}$$

$$e_c+N_ce_1-N_ce_2=i_cR_c \tag{80}$$

$$N_ci_c-i_{e2}=+K \tag{81}$$

$$N_ci_c+i_{e1}+N_si_{L1}-N_pi_{ac}=-K \tag{82}$$

Solving the foregoing equations yields

$$e_1=\frac{K(R_c/N_c^2-2AR_e)+AN_pR_ei_{ac}-e_c/N_c}{1+AB} \tag{83}$$

$$e_2=\frac{N_pR_ei_{ac}+Be_c/N_c-\pi\left(B\frac{R_c}{N_c^2}+2R_e\right)}{1+AB} \tag{84}$$

$$i_{L1}=\frac{1}{1+AB}\frac{N_s}{R_L}\left[K\left(\frac{R_c}{N_c^2}-2AR_e\right)+AN_pR_ei_{ac}+\frac{e_c}{N_c}\right] \tag{85}$$

$$i_c=\frac{K}{N_c}+\frac{e_2}{N_cR_e} \tag{86}$$

After saturation of the setting core,

the equivalent circuit is given by Fig. 7(A). The circuit equations are again identical to the original system relations with the exception of the additional constant in the current equation.

$$e_c - N_c e_0 = i_c R_c \quad (87)$$

$$N_c i_c - i_e = +K \quad (88)$$

$$e_0 = i_e R_e \quad (89)$$

$$e_0 = \frac{e_c/N_c}{A} - \frac{K}{A}\frac{R_c}{N_c^2} \quad (90)$$

$$i_c = \frac{e_c}{R_c} - \frac{e_c}{AR_c} + \frac{KR_c}{AN_c} = \frac{e_c}{AN_c^2R_e} + \frac{KR_e}{AN_c} \quad (91)$$

Refer again to equations 18, 19, and 20. Substitute into them respectively equations 84, 90, and 83. Once more remembering that the total sum of the flux changes during one complete cycle is zero, substitute the results of the above substitution into equation 21 to obtain:

$$\frac{B\dfrac{E_c}{N_c} + R_eN_pI_M - K\left(B\dfrac{R_c}{N_c^2} + 2R_e\right)}{1+AB} \times \int_0^{\theta(p)} d\theta + \frac{\dfrac{E_c}{N_c} - K\dfrac{R_c}{N_c^2}}{A}\int_{\theta(p)}^{\pi} d\theta$$

$$= -\frac{E_c/N_c + AR_eN_pI_M + K\left(\dfrac{R_c}{N_c^2} - 2AR_e\right)}{1+AB} \int_0^{\theta(p+1)} d\theta \quad (92)$$

Integrating, let $E_c/N_c = E_c'$ and $R_eN_pI_M = I_M'$ again to obtain a difference equation relating the p and $p+1$ cycles.

$$\left(A^2I_M' - AE_c' - 2A^2KR_e + AK\frac{R_c}{N_c^2}\right)\theta(p+1) - \left(AI_M' - E_c' - 2AKR_e + K\frac{R_c}{N_c^2}\right)\theta(p) = \pi(1+AB)\left(E_c' - K\frac{R_c}{N_c^2}\right) \quad (93)$$

Integrating equation 85 to obtain the average load current in a given half cycle:

$$I_L(p) = \frac{1}{1+AB}\frac{N_s}{R_L}\left[K\left(\frac{R_c}{N_c^2} - 2AR_e\right) + AI_M' - E_c'\right]\frac{1}{\pi}\int_0^{\theta(p)} d\theta$$

$$\theta(p) = \frac{\pi(1+AB)R_L}{N_s} \times \frac{I_L(p)}{\left[AI_M' - E_c' + K\left(\dfrac{R_c}{N_c^2} - 2AR_e\right)\right]} \quad (94)$$

$$\theta(p+1) = \frac{\pi(1+AB)R_L}{N_s} \times \frac{I_L(p+1)}{\left[AI_M' - E_c' + K\left(\dfrac{R_c}{N_c^2} - 2AR_e\right)\right]} \quad (95)$$

Substituting equations 94 and 95 into 93 obtains the difference equation for the load current:

$$I_L(p+1) - \frac{1}{A}I_L(p) = \frac{N_s}{AR_L}\left(\frac{E_c}{N_c} - K\frac{R_c}{N_c^2}\right) \quad (96)$$

This equation has exactly the same form as that of equation 29 which described the performance of the system with a perfect **B-H** characteristic. The dynamic performance of the system will therefore remain unchanged. The time constant, power gain, and figure of merit are described by equations 33, 34, and 35. However, since the right-hand side of the equation is a function of E_c minus a constant, the steady-state relation of I_L versus E_c must be modified.

$$I_L(\infty) = I_L(\infty+1)$$

$$I_L(\infty) = \frac{N_c^2N_sR_e}{R_cR_L}\left(\frac{E_c}{N_c} - K\frac{R_c}{N_c^2}\right) = \frac{N_sR_e}{R_L}(I_cN_c - K) \quad (97)$$

Equation 97 indicates that the steady-state load current will theoretically go to zero when

$$E_c = \frac{KR_c}{N_c} \text{ or } I_c = \frac{K}{N_c} \quad (98)$$

Consideration of the assumed **B-H** characteristic shows that the output will remain at zero for $E_c < KR_c/N_c$. The net result then of assuming a **B-H** characteristic with finite and constant hysteresis is to give the same steady-state I_L versus I_c characteristic that may be expected from the perfect core, except that all values of I_c have been displaced by the constant amount K/N_c.

Conclusions

The parallel loaded, full-wave magnetic amplifier has been explored. A self-saturated transformer, the transactor, has been developed. The salient features of its experimental behavior have been explained by the use of a series of elementary **B-H** core characteristics.

The transactor offers performance which in several respects is superior to that of the conventional series-loaded magnetic amplifier. The advantages are:

1. Improved transient and steady-state output stability for extreme variations of line, load, and frequency.
2. The quality of being fail-safe: a short circuit on the output will not overload it.
3. A rectangular, pulse-width modulated output wave with the following characteristics:
   a. It is ideal for driving stages of transistor power amplification.
   b. With rectification, it provides a superior source of power for precisely controlled d-c motors. The high, sharp pulse is ideal for overcoming the system "stiction" which prevails with low average voltages and consequently very slow speed.
4. An output voltage which is reducible to almost zero regardless of load impedance; the transactor can work into an open circuit.
5. Magnetic interstage coupling which is conveniently achieved; quiescent current effects are reduced.
6. Use of the half-wave transactor as a building block for new circuits including a full-wave phase-reversible amplifier and a variation of Patten's pulse stretch concept.[9] (These variations of the transactor will be described in Part II.)
7. Convenient temperature compensation: the analysis of the transactor, confirmed experimentally, shows the output to be a direct function of the core eddy-current resistance. Temperature compensation is therefore readily achieved by connecting appropriate temperature compensating resistors across a winding of each core of the transactor.
8. Multiple transactor operation from a single limiter. By proper proportioning of the primary windings a number of transactors may be fed in series from a single limiter without interaction of the outputs.

More exact correlation of the transactor's experimental and theoretical performance will require the use of more sophisticated **B-H** characteristics.[10,11] In addition, further consideration of additional second- or third-order effects such as interwinding capacitance, rectifier leakage, and after-saturation inductance will become necessary. With the use of lower-quality cores possessing less square **B-H** loops, the effects of after-saturation mutual inductance should become particularly noticeable.

References

1. Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load, **R. E. Anderson.** *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 76–87.

2. Saturating Transformer Reference Circuit, **W. G. Evans.** *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 610–16.

3. On the Mechanics of Self-Saturating Magnetic Amplifier Operation, **T. Kusada.** *Journal*, Institute of Polytechnics, Osaka City University, Osaka, Japan, Series E, vol. 3, Mar. 1956, pp. 103–17.

4. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955.

5. Analysis of Magnetic Amplifiers by the Use of Difference Equations, **P. R. Johannessen.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

6. Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions, **H. C. Bourne, Jr., D. Nitzan.** *Ibid.*, vol. 76, pt. I, Nov. 1957, pp. 646–55.

7. The Block-Diagram Method of Analysis.

APPLIED TO THE SATURABLE REACTOR, R. M. Hubbard. *Ibid.*, vol. 77, pt. I, Mar. 1958, pp. 57–65.

8. ANALYSIS OF THE FULL-WAVE MAGNETIC AMPLIFIER CIRCUITS CONSIDERING THE CHANGE OF THE WIDTH OF THE DYNAMIC HYSTERESIS LOOP, Tadashi Kikuchi. *Ibid.*, vol. 76, pt. I, July 1957, pp. 241–49.

9. THE PULSE STRETCH COUPLING CIRCUIT, H. W. Patton. *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 377–79.

10. CORE-RESET FUNCTIONS IN MAGNETIC AMPLIFIER ANALYSIS, PART I—A CORE-RESET FUNCTION, G. C. Feth. *Ibid.*, vol. 77, pt. I, 1958 (*Paper no. 57–95*).

11. CORE-RESET FUNCTIONS IN MAGNETIC AMPLIFIER ANALYSIS, PART II—DETERMINATION OF AMPLIFIER PERFORMANCE. *Ibid.* (*Paper no. 57-921*).

◆

Discussion

R. M. Hubbard (Boeing Airplane Company, Seattle, Wash.): The author is to be commended for his analysis of a type of magnetic amplifier circuit connection which has been seriously neglected in the past. The discusser has also been interested in this general type of connection and has found only one earlier reference[1] to the technique described in the paper for coupling the load to a magnetic amplifier. However, it is felt that the author in his analysis has overemphasized the uniqueness of this load coupling technique to the point of beclouding the fact that the transactor is not fundamentally different from the conventional self-saturating circuit. Basically the transactor may be classified as a self-saturating magnetic amplifier in which the conventional load is termed a "limiter" while the true load is transformer-coupled in shunt with the gate winding. The fact that the author has provided a unique type of limiter in the form of a series saturable reactor should not alter this statement. Since it has been this discusser's practice in the past to draw attention to the basic and extensive similarities between different classes of magnetic amplifier circuits by treating them with a single method of analysis,[2,3] he is inclined to believe that the author's conclusions about voltage gain, current gain, dynamic power gain, the effect of eddy-current resistance, etc., are identical to the conclusions one would reach in the treatment of a conventional self-saturating circuit. For example, assuming that the author's use of a linear eddy-current resistance R_e, to account for the finite excitation impedance of the core, is equivalent to the discusser's use of an average nonlinear resistance $\bar{R}$, one finds that transactor equations 31, 33, and 35 are equivalent to equations 5, 15, and 48 respectively, in the discusser's analysis of the simple full-wave self-saturating circuit.[2]

I am inclined to take exception to several of the author's statements relating to the representation of the magnetic core characteristics in terms of equivalent linear elements. First, as has been previously indicated, the conclusion that the transactor output is a direct function of eddy-current resistance R_e is also directly applicable to the simple self-saturating circuit. Therefore, if this fact in reality makes for convenient temperature compensation of the transactor, then the same is true of the conventional circuit and no advantage of one circuit over the other can be claimed on the strength of this point. More important, however, I do not feel that the representation of the equivalent eddy-current resistance as a linear resistance is justifiable in view of physical realities, in spite of the fact that such an assumption is essential to a straightforward analysis in terms of instantaneous values. This point is more clearly seen if one considers a transactor in which the excitation current is sinusoidal. Under such conditions, the flux-current loop would be substantially elliptical rather than rectangular. One is forced to conclude that an equivalent eddy-current resistance must be instantaneously nonlinear. In contrast, the beauty of a half-cycle average type of analysis lies in the fact that even though the eddy-current resistance be instantaneously nonlinear, the average value may be treated as a linear quantity. This is analogous to the fact that even though the functional relationship between the input and the output of a magnetic amplifier is discontinuous, this relationship in terms of half-cycle average values can be substantially linear. Thus, it is felt that the author's conclusion, in the section "Steady-State Performance" to the effect that the performance of the transactor will depend upon the existence of a constant value of eddy-current resistance, is unnecessarily restrictive and should in reality state that the performance is dependent on the existence of a constant half-cycle average value of eddy-current resistance.

Since, in practice, the excitation impedance of a square-loop core is substantially resistive both statically and dynamically, it is felt that the author's neglect of the static loop by considering only the eddy-current resistance does not conflict excessively with physical reality (except in the case of ferrites). However, it is felt that the author's "more sophisticated" **B-H** loop in which the eddy-current resistance is neglected, and the core excitation impedance represented by a linear inductance L_p, is in complete disagreement with physical reality. Even static measurements in which eddy-current losses are truly negligible reveal the square-loop core to be a dissipative element with very little energy storage. The representation of the core excitation requirements in terms of linear inductance implies that the core is substantially lossless. In fact, the author's assumption of a linear magnetizing inductance is incompatible with the corresponding **B-H** characteristic of Fig. 15 which is valid for a 100% lossy element. It is suspected that the exponential-like decay illustrated by Fig. 18 may be more realistically related to a natural rate of change of flux under conditions of constant current drive independent of R_e and thus independent of any L/R time constant. Investigations of this sort have been reported by Huhta[4] and others.

The problem of deriving a satisfactory equivalent circuit for the saturable transformer utilizing square-loop cores is a fascinating one. The work of Kikuchi and of Feth, (see references 8 and 10 respectively

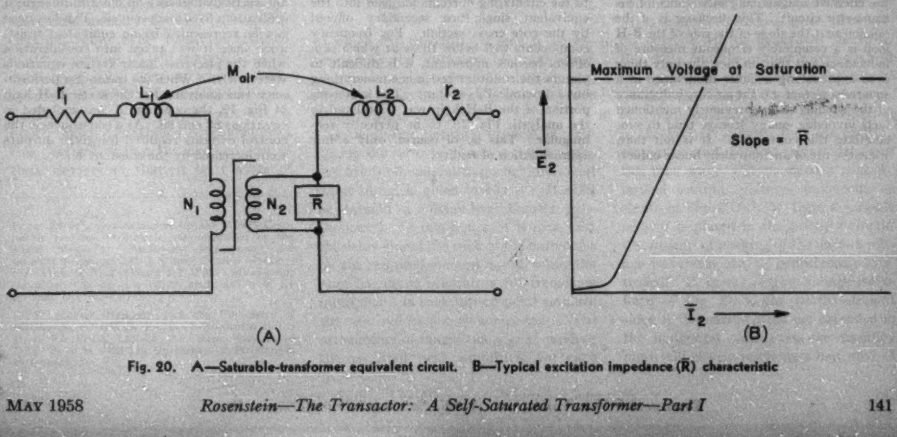


Fig. 20. A—Saturable-transformer equivalent circuit. B—Typical excitation impedance ($\bar{R}$) characteristic

of the paper) has shed considerable light on the dynamics of core operation with reference to certain restrictive operating conditions. However, nothing approaching a general analysis of core behavior which is not restricted by waveforms, impedance levels, etc., has yet appeared. I have found that this problem may be circumvented to a great extent by restricting the discussion to half-cycle average values. As a result, the excitation impedance may be represented as a nonlinear average resistance which is surprisingly independent of the waveforms of current and voltage involved.

An equivalent circuit for the saturable transformer, utilizing a square-loop core, which has been found compatible with a variety of experimental observations is presented in Fig. 20(A). The "ideal" transformer is an ideal saturating transformer which has infinite permeability in the unsaturated region and zero permeability in the saturated region. The linear inductances L_1 and L_2 correspond to the air-core inductances of the N_1 and the N_2 windings respectively, and the resistances r_1 and r_2 correspond to the winding resistances of the N_1 and the N_2 windings respectively. The quantity $\hat{R}$ is the nonlinear, half-cycle average excitation impedance of the N_2 winding. $\hat{R}$ may typically be determined empirically with a simple variable reset, half-wave test circuit. Typically, though $\hat{R}$ is nonlinear, it will be found to be characterized by two substantially linear regions, as illustrated in Fig. 20(B). The first is a low-impedance region corresponding to the threshold characteristic of square-loop cores in which very little flux change occurs. The second region is a high-impedance region encountered when the threshold ampere turns are exceeded. One limitation on this equivalent circuit is imposed by the fact that practical materials do not saturate as sharply as the "ideal" core, with the net result that the average voltage across $\hat{R}$ actually rises slightly as the core is driven into saturation.

It is commonly assumed that the inductance of a winding on a square-loop core has two distinct values: one associated with the high differential permeability corresponding to the slope of the side of the **B-H** loop and the other associated with the low permeability in saturation. The author has used the former value of permeability to calculate the effective magnetizing inductance of his transactor circuit. This discusser is of the opinion that the slope of the side of the **B-H** loop is a completely erroneous measure of inductance and that, in fact, the only thing even approaching a linear inductance in a square-loop reactor is the air-core inductance of the winding. The previously mentioned facts involving energy storage tend to substantiate this contention. It is felt that, while the use of an equivalent linear inductance and resistance may be justifiable in the representation of the excitation impedance of low-remanence transformer cores, a similar attempt to linearize such a distinctly nonlinear element as a square-loop core is doomed to failure.

REFERENCES

1. **MAGNETIC AMPLIFIER, R. W. Roberts.** *U.S. Patent no. 2,795,752*, June 1957.

2. **MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, R. M. Hubbard.** *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 578-88.

3. **THE BLOCK-DIAGRAM METHOD OF ANALYSIS APPLIED TO THE SATURABLE-REACTOR, R. M. Hubbard.** *Ibid.*, vol. 77, pt. I, Mar. 1958, pp. 57–65.

4. **FLUX RESETTING CHARACTERISTICS OF SEVERAL MAGNETIC MATERIALS, Hoobert Huhta.** *Ibid.*, vol. 73, pt. I, May 1954, pp. 111–14.

Allen B. Rosenstein: The author wishes to thank Mr. Hubbard for his thorough review. Mr. Hubbard's own works were valuable sources of information and guidance.

The author agrees that the transactor is one of a family of self-saturating magnetic amplification devices. In each case a large impedance is reflected into the supply circuit during mode II while the control voltage effectively divides itself between the reflected resistances and the series resistance R_c. During mode I, all circuits are coupled. Under these circumstances it is indeed natural to expect the parallel and series loaded amplifiers to possess identical transfer features. However, it should be noted that the use of a resistance limiter would have modified the gain equations of the transactor since the resistance of the limiter would be reflected in parallel with the load. It was not the intention to disguise the true parents of the transactor. It was hoped that the section "Parallel Output" along with Fig. 3 would clearly define the relationship.

The author is not certain whether the differences in the nature of the eddy-current resistance are conceptual, physical, or semantic. It was intended to separate the low-frequency characteristics of the core into its hysteresis and eddy-current components with the eddy currents being restricted to the circulating currents induced into the equivalent single-turn secondary offered by the core cross section. For frequency components well below those at which skin effects become important, it is difficult to picture the resultant resistance as anything but a dynamically constant. The hysteresis portion of the **B-H** loop was postulated in the analysis, Fig. 19, to be perfectly rectangular. This is, of course, only a fair approximation of reality.

An effort was made to determine theoretically and experimentally the mechanism of output voltage variation with change in limiter current. With a core postulated to possess only a constant eddy-current resistance and a rectangular static hysteresis loop, the output is completely independent of the limiter current. Since experimentally the output was found to vary somewhat with limiter current, one is lead to investigate further the possibilities of finite permeability in either the saturated or unsaturated region. A series of experiments was conducted to determine the more likely suspect.

Referring to the limiter voltage e_{lim} in Fig. 10, it may be seen that by biasing the limiter high enough, the limiter voltage will go to zero during the gating interval. This means that the x shown in Fig. 10 will be zero during the entire interval and all of the supply voltage will appear across the output. Under these circumstances, the limiter will not control the current during the gating interval, but will limit the load current only after the transactor has saturated. This means that changes in limiter current will only cause current changes through the aftersaturation inductance. Under these circumstances it was found that substantial changes in limiter current had no effect upon the transactor output. However, if with x still equal to zero, the a-c supply voltage was varied causing a change in the instantaneous magnitude of current flow through the transactor during the gating interval, then the same changes in output would appear as when $x \neq 0$ and the limiter current was varied. From this experiment and the behavior of the transactor reset voltage shown in Fig. 18 it was concluded that the aftersaturation inductance of the experimental transactor was not as important as the finite unsaturated permeability or magnetizing inductance.

The **B-H** characteristics with hysteresis and finite unsaturated permeability of Fig. 15 were assumed to investigate analytically these conclusions. To reduce the labor of the rather cumbersome equations generated by the inclusion of inductance in the equivalent circuits, it was decided not to consider the eddy-current resistance. As the discusser points out, this does not mean that the cores are lossless. However, since the static hysteresis loop has a finite area, it will cause a fixed loss per cycle. This loss need not be represented by an equivalent resistance since it was taken into consideration when the piecewise linear system equations were written. When the transactor performance was analyzed for the static **B-H** loop of Fig. 19, the constant K was included in equations 82 and 88. As a consequence, the control currents required for given outputs were increased by the constant K/N_c.

Vol.
Pt.
195

The Design of Function Generators Using Short-Time Memory Devices and Nonlinear Elements

A. W. REVAY
ASSOCIATE MEMBER AIEE

D. J. FORD
ASSOCIATE MEMBER AIEE

THE ANALOG COMPUTER FIELD is frequently concerned with the generation of periodic nonsinusoidal functions of time. At present this waveform generation is frequently performed with biased diode networks, curve followers, and synchronous switches. The main purpose of this paper is to give the theoretical design and analysis of nonsinusoidal function generators which utilize delay-type or short-time memory devices and certain nonlinear elements in their structure. The different types of function generators to be discussed have the common property that the output waveform may be controlled easily and may be made to approximate various desired shapes to high degrees of accuracy.

This paper is meant to be exploratory, in a theoretical sense only, of a sizable number of different methods of waveform generation using short-time memory devices and certain nonlinear elements. It is hoped at a later date to submit further theoretical work together with the experimental verification and quantitative evaluation of some of these methods.

Throughout this paper the delay-type device will be assumed ideal, i.e., the attenuation along the device will be assumed zero, and the phase shift will be assumed linear throughout the frequency range of interest. The increased use of short-time memory devices, particularly electromagnetic delay lines, has resulted in the development and commercial production of delay devices which closely approximate these ideal conditions over a significant range of frequencies. It is hoped that the continued development of this field will reduce the size and cost of these devices so that at least some of the methods suggested in this paper will find economical application.

Delay-Line Fourier Synthesizer

In all of the work which follows, the delay-type or short-time memory device will be assumed to be an electromagnetic delay line having the "ideal" characteristics over the indicated frequency range.

A large class of continuous periodic waveforms of engineering significance may be analyzed by Fourier series methods. A typical expression for such a waveform is

$$y(t)=C_0+C_1\sin(\omega t-\phi_1)+C_2\sin(2\omega t-\phi_2)+\ldots+C_n\sin(n\omega t-\phi_n)+\ldots \quad (1)$$

where the constants C_0, C_1 ... are obtainable from various integrals of the given function, $y(t)$, and from sine and cosine functions by well-known mathematical methods. Although the series in equation 1 has an infinite number of terms, there exist many important periodic waveforms for which relatively few terms give a close approximation to the given nonsinusoidal waveform. To a high degree of accuracy one of these periodic waveforms may be represented by

$$y(t)=C_0+\sum_{k=1}^{m}C_k\sin(k\omega t-\phi_k) \quad (2)$$

in which the waveform is seen to be expressed by the finite sum of a number of sine waves of different frequencies, different peak amplitudes, and different phase shifts or phase delays.

A delay-line network capable of producing the series of equation 2 when subject to a fundamental-frequency sinusoidal input signal is given in Fig. 1. It shall be termed a "delay-line Fourier synthesizer." In this figure, it is seen that the delay-line shifts each of the harmonics by the required amount ϕ_K/K, while the pure harmonic generators (or frequency multipliers) in each tap-off point produce the required frequency component. The attenuators of magnitude C_K/M produce the required peak amplitude for each frequency component. It is noted that each phase shift required of the delay line in only (ϕ_K/K) rather than ϕ_K itself. In particular cases this might be of considerable advantage in that the total magnitude of phase shift required for the delay line in terms of the fundamental frequency, is 180 degrees maximum, and might be considerably less. The fact that all of the phase shift angles obtainable from the delay line are negative poses no difficulty when it is recognized that, on a steady-state basis, any calculated positive angle ϕ_K may be replaced by $(\phi_K-360$ degrees). This latter angle will be negative since ϕ_K is by definition less than 360 degrees. The substitution of $(\phi_K-360$ degrees) for ϕ_K in the phase shift portion of a component of equation 2 does not alter the resulting steady-state waveshape.

It is noted that the main components of the delay-line Fourier synthesizer of Fig. 1 are the harmonic generators required for each tap-off point along the line. The following three sections describe three different ways in which these harmonic generators might be designed.

MAGNETIC-CORE HARMONIC GENERATOR

The magnetic-core harmonic generator presents a method of controlling a voltage waveshape by means of the nonlinear characteristics of a magnetic core. When several magnetic cores are used in conjunction with a delay line, it is possible to produce a harmonic component as needed for the Fourier synthesizer and indicated in equation 2. The input signal to the magnetic-core harmonic generator is assumed to be a sine wave of the fundamental frequency required in the waveform synthesis.

If a sinusoidal voltage e_g is applied to the input of the circuit shown in Fig. 2(A), and if the primary circuit resistance (R_g+R_1) is assumed to be negligible, the exciting current will be peaked as shown in Fig. 2(B), but the output waveform will be essentially sinusoidal as indicated.[1] On the other hand, if R_g and R_1 are large, or if additional large resistance is inserted in the primary circuit, then the exciting current would become sinusoidal, and the flux wave will produce a sharply peaked output voltage waveform as shown in Fig. 2(C).[2] If, then, a variable resistor is placed in the primary circuit, the amount of peaking of the output voltage waveform can be continuously controlled. If, in the output voltage waveform of Fig. 2(C), the nearly straight sides of the narrow peak are extended to the horizontal time axis, the resulting pulse will closely resemble a half-wave of

---

Paper 58-247, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing December 5, 1957.

A. W. REVAY, formerly with the University of Pittsburgh, Pittsburgh, Pa., is a lieutenant with the U. S. Air Force, Lincoln AF Base, Neb., and D. J. FORD is with the University of Pittsburgh, Pittsburgh, Pa.

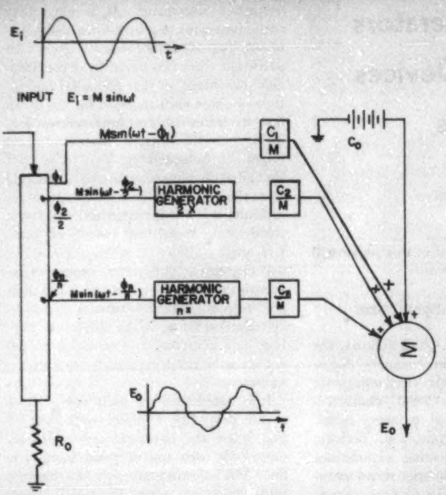


Fig. 1. Fourier synthesizer using electromagnetic delay-line to shift harmonics. $E_0 = C_0 + C_1 \sin(\omega t - \phi_1) + C_2 \sin(2\omega t - \phi_2) + C_3 \sin(3\omega t - \phi_3) \ldots . C_n \sin(n\omega t - \phi_n)$

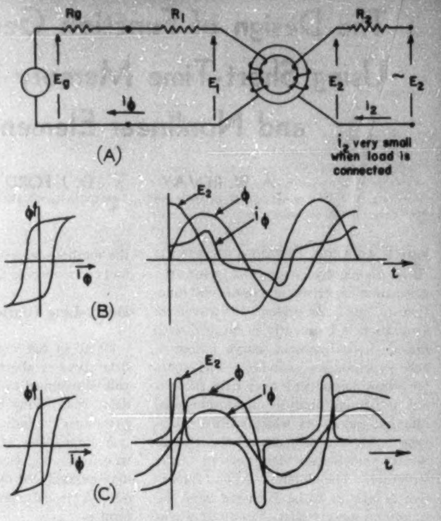


Fig. 2. Magnetic-core harmonic generator unit

A—Circuit diagram

B—Output waveform with sinusoidal voltage input and low primary resistance

C—Output waveform with sinusoidal voltage input and high primary resistance

sine wave of higher frequency than the fundamental. If the pulse width is defined as the angle measured between the intersections of the extensions of the nearly straight sides of the pulse and the horizontal axis, the pulse width may be varied from 180 degrees to some value 180 degrees $/S$, where S is limited by the hysteresis loop of the iron and the peak input voltage. When S is adjusted to a value 2.00, the two pulses appearing in an output voltage cycle will approximate the positive and negative half-cycles of the second harmonic. In general, for integer values of S, say K, the pulse portions of the output waveform will approximate the half-cycles of the Kth harmonic; see Fig. 3.

A true, relatively pure Kth harmonic sinusoidal waveform may be produced by connecting either K or $2K$ iron cores in delay-line networks as shown in Fig. 4. The second and third harmonics are used as examples in these figures, but the general configurations may be directly extended to other odd and even harmonics. From these figures, it is seen that the iron-core circuits are connected along the delay lines in such a fashion that their input signals are displaced from one another by 180 degrees$/K$, the angles being measured with respect to the fundamental frequency component. By summing the successively displaced pulsed waveforms, the resultant relatively pure harmonic output is produced. The small amplitude sections of the individual core outputs which link the positive to the negative pulses tend to cancel out in the summation used to produce the output signal.

Experimentally, the value of primary resistance needed could be determined by adjusting the variable resistor indicated in the primary circuit until the desired peaking of the waveform is produced. The variable resistor could then be replaced by a fixed resistance element. Because of the small size of the magnetic cores, the physical size of this magnetic-core harmonic generator would be quite small. Because only the fundamental sinusoidal waveform input voltage passes through the delay line, many of these harmonic generators could be connected to the same delay line. If there were sufficient taps on the delay line, or if a mechanism for continuous adjustment of the delay time were available, it would be possible to shift each of these harmonic generators as a complete unit by the proper phase amount ϕ_K, as determined by equation 2.

In cases where the iron-core hysteresis loop limits the number of harmonics obtainable, one of these obtainable harmonics may be connected as the input to an auxiliary delay line. The magnetic-core harmonic generators connected at output taps of the auxiliary delay line will give harmonics of the input harmonic.

The number of harmonics might also be increased by placing a series capacitor in the primary circuit to "tune out" the inductive effects and permit an exact sinusoidal current waveform. However, even without these extra measures, the square-hysteresis-loop magnetic cores commercially available will provide a sufficiently high number of harmonics for most Fourier synthesis applications.

It might seem desirable to utilize the alternate method of using one magnetic core to obtain the desired pulse width and then to shift these pulses by a delay line. The resultant waveform would then be obtained by summing the shifted pulses in the usual manner. Although this might be satisfactory for certain frequency ranges of operation, it is thought that the high harmonic content of the pulses would require a delay line

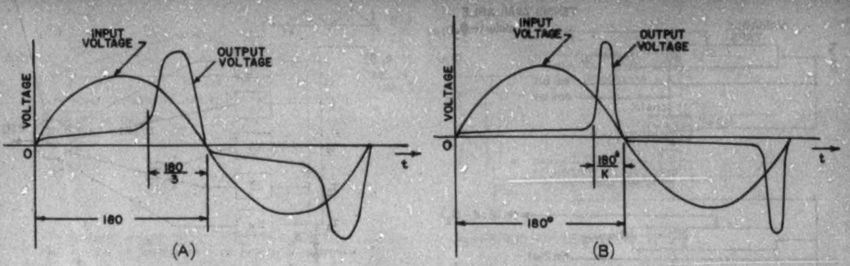


Fig. 3. Output waveform of magnetic-core harmonic generator unit showing peaking: A—to form a portion of third harmonic. B—to form a portion of Kth harmonic

having a relatively much higher cutoff frequency than the network just considered. This system would require either one delay line for each Fourier series component, or a complicated set of filters to receive the respective pulses. The first case would seem to be economically unsound compared to the system discussed because delay lines are much more costly than iron cores. The second method would probably be economically unsound also, since the complicated set of filters could just as easily be used with the original distorted wave. This would eliminate entirely the need for the delay-line mechanism.

Nonlinear-Element Harmonic Generator

A network or device which possesses a specific type of nonlinear volt-ampere characteristic may serve as a harmonic generator. To illustrate, consider a solid-state diode or device whose volt-ampere characteristic is describable by

$$i = Ke^{a} \tag{3}$$

If a pure fundamental-frequency sinusoidal voltage is applied to this device with the loading effects adjusted so that equation 3 gives the dynamic volt-ampere characteristic with *a* equal to two, a second harmonic component of current will be generated as indicated by

$$i = K(E_m \sin \omega t)^2 \tag{4}$$

$$= \frac{KE_m}{2} - \frac{KE_m}{2} \cos 2\omega t \tag{5}$$

The d-c component would have to be balanced out, or blocked by capacitive coupling, or possibly used to form the d-c term in the Fourier series of equation 2.

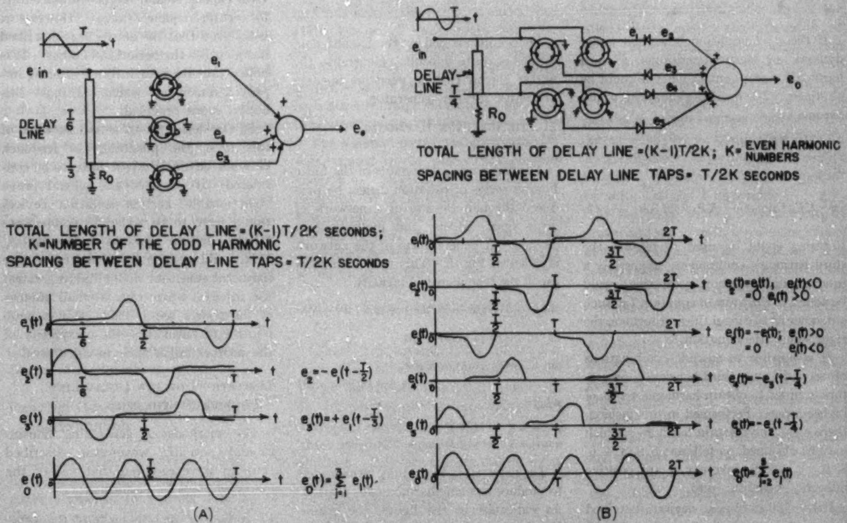
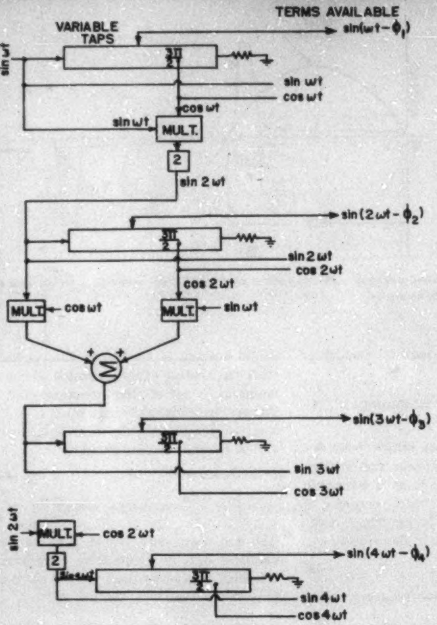

T
t
e in
e₁
DELAY LINE
T/6
T/3
R₀
e₂
e₃
eₒ
TOTAL LENGTH OF DELAY LINE = (K-1)T/2K SECONDS;
K = NUMBER OF THE ODD HARMONIC
SPACING BETWEEN DELAY LINE TAPS = T/2K SECONDS
e₁(t)
e₂(t)
e₃(t)
eₒ(t)
e₂ = −e₁(t − T/2)
e₃(t) = +e₁(t − T/3)
eₒ(t) = Σ_{j=1}^{3} eⱼ(t)
(A)
T
t
e in
DELAY LINE
T/4
R₀
e₁
e₂
e₃
e₄
e₅
eₒ
TOTAL LENGTH OF DELAY LINE = (K-1)T/2K; K = EVEN HARMONIC NUMBERS
SPACING BETWEEN DELAY LINE TAPS = T/2K SECONDS
e(t)
T/2
T
3T/2
2T
e₂(t) = e(t); e(t) < 0
= 0 e(t) > 0
e₃(t) = −e(t); e(t) > 0
= 0 e(t) < 0
e₄(t) = −e₃(t − T/4)
e₅(t) = −e₂(t − T/4)
eₒ(t) = Σ_{j=2}^{5} eⱼ(t)
(B)

Fig. 4. Magnetic-core harmonic generator network

A—For third-harmonic generation. General network structure may be extended for generation of any odd harmonic
B—For second-harmonic generation. General network structure may be extended for generation of any even harmonic



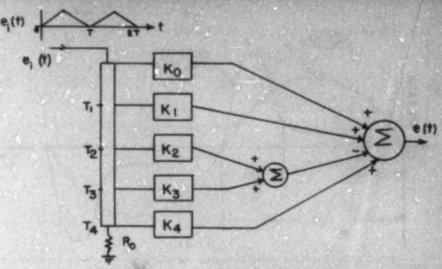


Fig. 6. Typical network for constant coefficient slope function generator which utilizes triangular input signal

Output signal $e_0(t)=K_0e_1(t)+K_1e_1(t-T_1)-K_2e_1(t-T_2)-K_3e_1(t-T_3)+K_4e_1(t-T_4)$

Fig. 5 (left). Multiplier type of waveform generator

If the a of equation 3 were 3 under dynamic or loaded conditions, a third harmonic component of current could be obtained. The resulting equations would become

$$i=Ke^3=KE_m{}^3\sin^3\omega t \qquad (6)$$

or

$$i=\frac{3}{4}KE_m{}^3\sin\omega t-\frac{1}{4}KE_m{}^3\sin 3\omega t \qquad (7)$$

Filtering could be used to remove the third harmonic component, or possibly a fundamental-frequency component could be added to the wave of equation 7 in such a manner as to cancel the first term on the right side of the equation.

It is possible to cascade combinations of second- and third-harmonic generators in order to obtain harmonic terms of higher order. Performed in this fashion, there exist only certain harmonics which may be obtained, as follows: 1, 2, 3, 4, 6, 8, 9, ... (2^x3^y), where x, y are positive integers, including zero.

Individual nonlinear devices described by equation 3 with a greater than 5 usually do not provide a sizable output signal amplitude. For some devices, the accurate control of the operating point conditions and/or the loading conditions is quite difficult. Generally, the use of silicon carbide varistors appears attractive for this generator.

MULTIPLIER-TYPE HARMONIC GENERATOR

The harmonic components of the finite Fourier series of equation 2 may be produced through the use of a network of delay lines, adders, and multiplication devices. A general form of the network is given in Fig. 5. This network makes use of the trigonometric identity

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta \qquad (8)$$

or, in more general form,

$$\sin K\omega t=\sin n\omega t\cos m\omega t+\cos n\omega t\sin m\omega t \qquad (9)$$

where

$n+m=K$ for K odd
$n=m=K/2$ for K even

The network of Fig. 5 may be extended to produce as many harmonics as desired. As indicated in the figure, the cosine function is readily obtainable at the $3\pi/2$ tap, or negatively at the $\pi/2$ tap. The phase angle of the nth harmonic may be varied if a variable delay-line tap is provided at the output. By choosing the proper phase shift and peak amplitude for each harmonic component, a given complex wave may be synthesized through the addition of the harmonic components obtained from this network.

At present, objection might be made to the frequent use of multipliers because in the present state of the art they are both expensive and relatively inaccurate for certain frequency ranges. However, it is believed that the network as described has no other theoretical difficulties. It is hoped that the near future will bring improved multipliers which will make this design more practical.

In the design of an actual network of this type, the possibility of feedback between channels might have to be considered. If the feedback should prove objectionable, certain isolation devices would have to be added to compensate for the parasitic elements causing it.

It would seem that this multiplier-type harmonic generator would find its greatest use in cases where only a small number of harmonics are needed. For a large number of harmonics, the complexity of the network might make its use unwise.

LIMITATIONS OF THE DELAY-LINE FOURIER SYNTHESIZER

The methods of generating nonsinusoidal periodic waveforms described thus far have required that one of the following steps be performed:

1. A Fourier analysis made of the waveform to determine the correct amplitudes and phase angles of the significant harmonics.
2. A trial and error experimental method

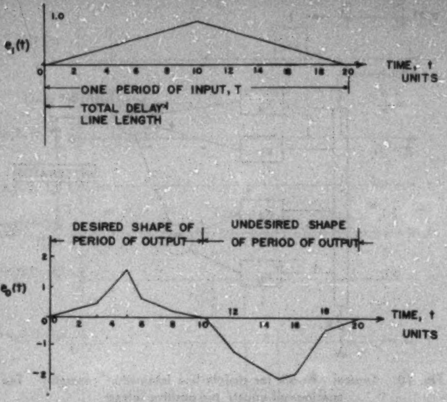


Fig. 7. Typical waveforms for constant coefficient slope function generator which utilizes triangular input, using symbols of Fig. 6: $T_1=3$, $T_2=5$, $K_0=1$, $K_1=1.2$, $K_2=4$, $T_3=6$, $T_4=8$, $K_3=1.5$, $K_4=3$. Time units are nondimensional, and equal to 20 (actual time/base time)

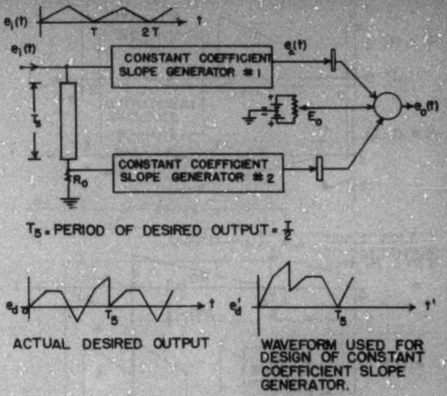

$e_i(t)$
T
2T
t
$e_i(t)$
CONSTANT COEFFICIENT SLOPE GENERATOR #1
$e_2(t)$
E_0
$e_0(t)$
T_5
R_0
CONSTANT COEFFICIENT SLOPE GENERATOR #2
T_5 = PERIOD OF DESIRED OUTPUT = $\frac{T}{2}$
e_d
T_5
t
ACTUAL DESIRED OUTPUT
e_d'
T_5
t'
WAVEFORM USED FOR DESIGN OF CONSTANT COEFFICIENT SLOPE GENERATOR.

Fig. 8. Typical network for composite restrictive slope function generator which utilizes triangular input signal

of amplitude and phase angle adjustment made, in order to fit the required waveform.

Both methods are usually time-consuming although necessary for the design of the foregoing networks.

The accuracy with which the resultant synthesized waveform approximates the ideal model is determined by the accuracy with which the magnitude and phase of the harmonic components are determined and by the number of harmonic components used to make up the resultant waveform. The number of harmonics required for a given accuracy of approximation is a function of the waveform to be synthesized. In general, greater accuracy is obtained by increasing the number of harmonic components used, but this procedure also increases the cost and the complexity of the generator. In a given case a compromise between cost and accuracy would have to be met. The method of waveform generation described in the next section does not require the time-consuming preliminary work indicated in the foregoing.

Function Generation by Waveform Approximation

A given periodic function of time may be approximated by fitting generated curve segments to various sections of the given curve. The two main classifications of curve fitting or waveform approximation which will be considered here are: 1. straight-line or first-degree approximation, and 2. curved-line or higher-degree approximation. The higher-degree approximation can give the greater accuracy for a given number of segments, if an effective means of controlling the segment curvature is available.

This section will describe several methods using short-time memory devices by which waveforms may be approximated by a straight-line segment method and by a higher-degree polynomial segment method. Both methods have the advantage of flexibility and ease of control.

The Kallman transversal filter network will be used in the proposed waveform generators.[8] A given set of networks will be connected at the delay line taps of the transversal network in order to give the over-all network various transfer characteristics as the input signal proceeds along the delay line.

STRAIGHT-LINE APPROXIMATION METHOD

There are several types of similar transversal networks which may be used under somewhat restrictive conditions for the purpose of function forming, with the use of straight-line segments. One of these, originally suggested by Sze, is given in Fig. 6. It is seen that the output signal $e_0(t)$ of Fig. 6 may be written in terms of the input signal $e_i(t)$ as

$$e_0(t)=K_0e_i(t)+K_1e_i(t-T_1)-K_2e_i(t-T_2)-K_3e_i(t-T_3)+K_4e_i(t-T_4) \quad (10)$$

In the general case, the sign preceding any K_i can be made plus or minus by connection to the proper summing device indicated in the figure.

One possible type of waveform input for this network is the triangular type shown at the top of Fig. 6. A typical set of input and output waveforms for this network is given in Fig. 7, a study of which will reveal that the control of the output waveshape for the interval $0<t<T/2$ is performed readily by manipulation of the K_j and the connections to the two summing devices in the network of Fig. 6. However, once this portion of the wave is specified, the later portion of the output wave (i.e., for the interval $T/2<t<T$) is completely determined, and cannot be altered without changing the output wave for the first half-period.

If the wave to be generated can be drawn so that its minimum value occurs at the beginning of the period, and thus correspondingly at the extreme end of the period, this difficulty can frequently be overcome. The composite restrictive slope function generator shown in Fig. 8 will then provide the desired output waveshape, provided that the output of, say, generator 1 is entirely negative for the interval $T/2<t<T$. The rectifiers in the network will completely block the extraneous output from this generator during this interval, while allowing the desired output from generator 2 to pass to the output terminals. The direct voltage E_0 may be adjusted to restore the orginal d-c level in the generated wave. Conceivably, there exist such waveforms that the adjustment of the coefficients to produce the e_d' waveform of Fig. 8 will

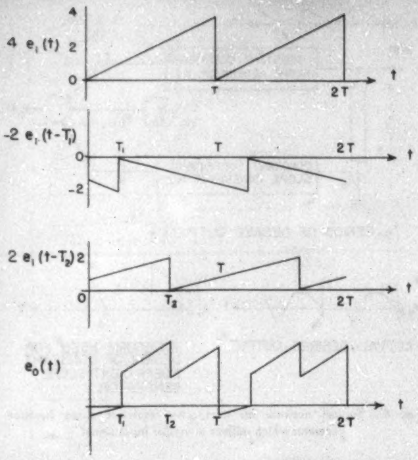


Fig. 9. Typical waveforms for constant coefficient slope function generator which utilizes saw-tooth input. The output $e_0(t)=4e_1(t)-2e_1(t-T)+2e_1(t-T)$, T = period of input $e_1(t)$

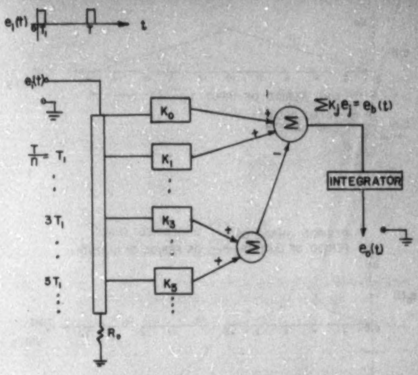


Fig. 10. Typical network for straight-line integration generator. Tap spacings all equal; n = positive integer

not provide an entirely negative output from the individual generators every alternate half-cycle of the input wave. In this case, the network would deliver an erroneous output every other full cycle of the output wave. Thus the composite-restrictive-slope function generator of Fig. 8 will perform satisfactorily for only a restricted class of desired output waveshapes.

One method which might appear to be a solution to the difficulty imposed on the function generator of Fig. 6 is to change the input to a saw-tooth shape so that the uncontrollable "second" half-cycle of the input signal is eliminated. A set of typical waveforms for this type of network is given in Fig. 9. Although this generated output signal might conceivably approximate a desired shape, one may imagine many shapes for which the sudden vertical changes in amplitude at the end of the component waveform periods would be intolerable.

If the saw-tooth input wave utilized in Fig. 9 were replaced by a saw-tooth pulse, the duration of which was small with respect to the period T of the wave, the resulting network configuration may be made to perform very satisfactorily for the generation of virtually all important periodic functions. Because this network and design procedure is a special case of the general polynomial approximation method, to be discussed in the next section, it will not be examined here.

The straight-line integration generator of Fig. 10 illustrates another slight variation of the basic tranversal network. When the tap spacings are all made equal to the square input pulse duration, there exists only one component of the K_jE_j sum which has a nonzero value at any one instant of time, illustrated by the typical waveforms shown in Fig. 11. This presents an appreciable advantage for the final experimental adjustment in that each coefficient clearly controls one section of the output curve exclusively. For the purposes of the general study here, the integrator is assumed to be ideal.

If the input signal to the straight-line integration generator of Fig. 10 were a conventional square wave with equal-duration ON and OFF intervals, the resulting performance would be imperfect every second half-cycle of the desired output. The basic reason for this is the lack of complete control of the component waveshapes throughout a full cycle of the input. This difficulty is in reality just a slightly different form of the difficulty associated with the constant coefficient slope function generator of Fig. 6.

POLYNOMIAL FUNCTION APPROXIMATION

It is assumed that most single-valued continuous periodic functions of engineering interest may be approximated over reasonable intervals of the independent variable by polynomial functions. The method utilizing this fact concerns itself with the discrete generation of polynomials to fit the given periodic function over prescribed portions of the fundamental period of the function. The resultant generated periodic function will then be the sum of these various polynomial functions, each valid over a different portion of the fundamental period of the resultant desired function. For the sake of concreteness, it will be assumed that each segment of the resultant desired function can be approximated closely by a fourth-degree polynomial. The first network to be designed is then a fourth-degree polynomial segment generator (PSG).

A general fourth-degree polynomial function of time may be expressed as

$$e_0(t)=A_0+A_1t+A_2t^2+A_3t^3+A_4t^4 \qquad (11)$$

in which the coefficients A_0, A_1, A_2, A_3, and A_4 are positive or negative real numbers. If it is possible to vary or adjust the A_j coefficients, then $e_0(t)$ may be made to approximate any real fourth-degree polynomial as needed for the particular segment under consideration. It is important to be able to obtain negative as well as positive A_j coefficients, for a restriction to positive values only would seriously hamper the type of representation possible.

A nonlinear resistance network described by Surber, Jr.,[4] and shown in Fig. 12 is able to produce an output voltage, defined by the equation

$$\begin{aligned} e_d(t) &= +K_a\left|e_i^a(t)\right|, \ e_i(t)\geq 0 \\ &= -K_a\left|e_i^a(t)\right|, \ e_i(t)<0 \end{aligned} \qquad (12)$$

The exponent a can usually be varied throughout the interval from one to four by the adjustment of r_3 and r_4 of Fig. 12.

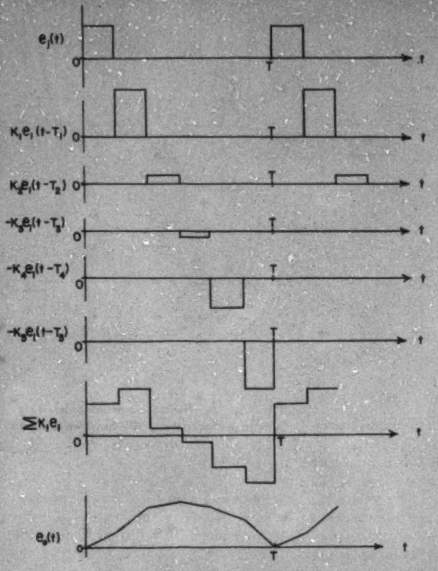


Fig. 11. Typical waveforms for straight-line integration generator. The output
$e_0(t)=\int_0^t \Sigma K_i e_i dt$

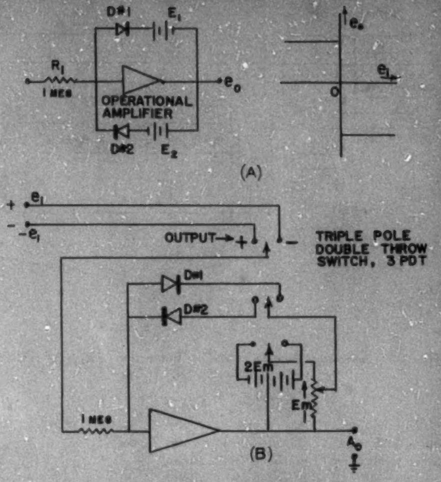


Fig. 13. A_0 term circuit for use in the PSG

A—Series output limiter with zero feedback resistance and its output-input characteristic

B—Adaptation to obtain direct voltage pulse only while PSG has an input signal

The double L-network of this figure contains two nonlinear Thyrite resistors as series elements, in addition to the two adjustable linear shunt resistance elements. Networks of the type illustrated in Fig. 12 would then be able to produce all of the terms on the right-hand side of equation 11, except the constant term A_0, provided the input is a linear function of time.

The insertion of the constant term A_0 into the network cannot be performed as simply as might be initially assumed. It is recognized that the proposed design utilizes several PSGs, each operating in turn, to form the segments that will then be joined to form the over-all desired waveform. At any instant, only one PSG will operate, so that it must have a proper A_0 term to join its boundaries with those of the preceding and succeeding segments. Using an average d-c component for all of the PSG units, or having a d-c term from one PSG present while another is operating would not permit the proper joining of the segments. Thus each PSG must have its proper A_0 term as an output, when and only when its segment is being contributed to the over-all waveform. To obtain the desired d-c output for each PSG, a special d-c generating device is needed which generates its output only when a nonzero input signal exists at the input to the PSG.

Fig. 13 shows the schematic diagram of a modified series output limiter which has the required properties for generation of the A_0 term. The output-input voltage characteristic indicates that the

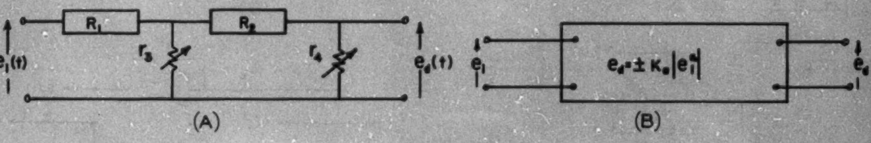


Fig. 12. Nonlinear double-L network

A—R_1,R_2 = nonlinear resistors
r_3,r_4 = variable linear resistors
$e_d(t)=+K_a|e_1^a(t)| \quad e_1(t)\geq 0$
$=-K_a|e_1^a(t)| \quad e_1(t)<0$
where a may usually be changed throughout the range: $1\leq a\leq 4$ by adjusting r_3,r_4
B—Representation of network in A

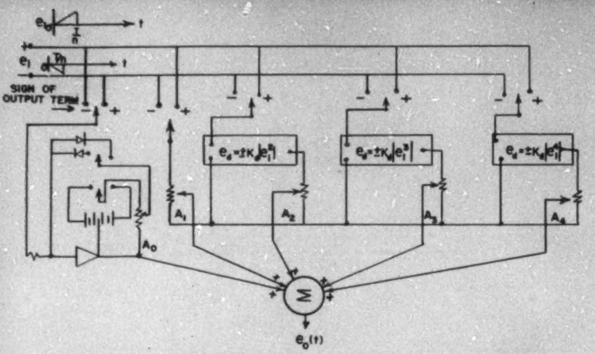


Fig. 14. Schematic diagram of PSG. The output $e_0(t)=A_0+A_1t+A_2t^2+A_3t^3+A_4t^4$ for $0<t<T/n$

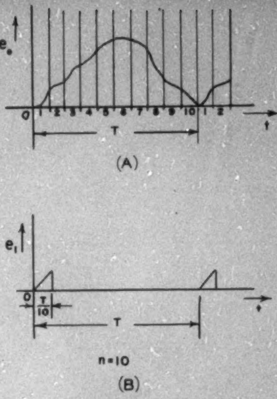


Fig. 16. A—Periodic nonsinusoidal waveform to be generated. B—Input waveform to delay line

output is constant and positive, or constant and negative, if the input has any nonzero negative or nonzero positive value respectively. For a zero input signal, this device delivers zero output. The more elaborate diagram on the lower portion of Fig. 13 indicates a form of the modified series output limiter which is compatible with the basic structure of the PSG.

The diagram of the complete PSG is given in Fig. 14. From this, it is seen that the output signal $e_0(t)$, is equal to the sum of the five terms constructed from the linearly increasing input. For the interval $0<t<T/n$, the output signal $e_0(t)$ is seen to be equal to the sum of the five terms as given in equation 11, because the positive and negative input signals are linearly increasing functions of time. The magnitude of the coefficients in this equation are readily obtained by adjustment of the potentiometers, while the sign associated with the coefficients is established by switching the respective term network input to the proper input channel.

The PSG of Fig. 14 is able to reproduce the general fourth-degree polynomial to a degree of exactness limited only by the linearity of the input and the exactness with which the coefficient can be established in the nonlinear networks. The fourth-degree polynomial is thus realized for the time segment $0<t<T/n$ indicated in the figure or, more particularly, for any interval of time during which the positive and negative input signals maintain a linearly increasing waveform.

There are many known ways by which a linearly increasing waveform may be established for a finite interval of time. In the generation of the final composite periodic waveform, it will be found convenient to have an input signal $e_1(t)$ which increases linearly with time for a finite interval, after which it is reduced abruptly to zero. This saw-tooth pulse is to be followed by a zero amplitude portion equal to $(n-1)T/n$ seconds, where T is the period of the wave to be formed, and n is the number of segments used in the approximation. By this mechanism, one PSG is able to act independently of the others in the complete waveform approximation circuit.

For the operation of the PSG, it is necessary to generate a train of positive-going triangular pulses together with the negative of this waveform. If a series of alternately positive and negative triangular pulses are available (such as might be generated from a sinusoidal source), the delay-line network of Fig. 15(A) will provide the required two sets of signals for the generator. If a series of exclusively positive triangular pulses is available, then the conventional operational amplifier inverter network of Fig. 15(B) may be used. These methods together with other known methods may

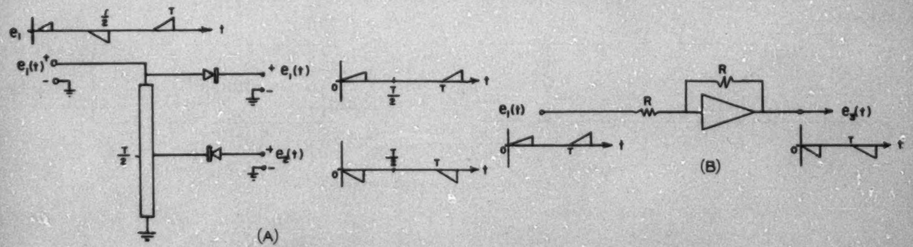


Fig. 15. Two methods for obtaining positive and negative input signals for PSG

A—Delay-line method B—Operational amplifier method

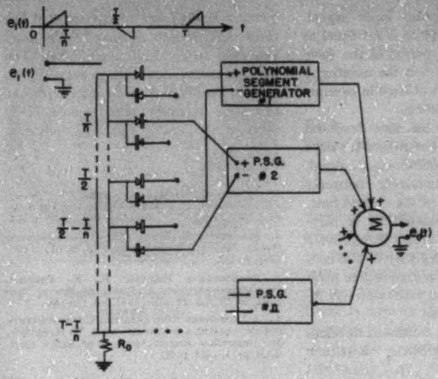


Fig. 17. Waveform approximation generator using n PSG and delay-line inverter network

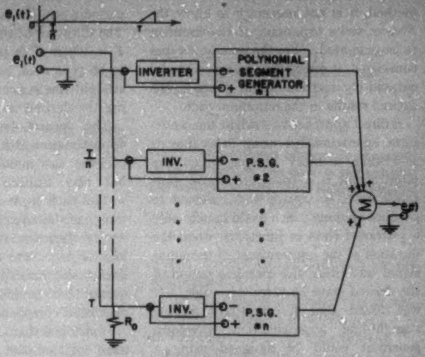


Fig. 18. Waveform approximation generator using n PSG and n inverters

be readily applied to the problem of producing the positive-going set of triangular pulses $e_i(t)$, together with its negative $[-e_i(t)]$.

The operation of the complete waveform approximation generator using n PSG may perhaps best be described in terms of a particular example. Suppose it is desired to reproduce a given periodic waveform to a very high degree of accuracy. Let this waveform as shown in Fig. 16(A) be divided into ten segments, and let each segment be approximated by a fourth-order polynomial function of time. This is equivalent to setting n, the number of waveform segments or the number of PSG equal to ten. The required duration of the triangular input pulse would then be $T/10$ where T is the period of the waveform to be approximated. One form of waveform approximation generator with the use of a delay-line inverter network which would produce this waveform is given in Fig. 17, in which n would be set equal to ten. An examination of this network shows that the PSG 1 delivers the entire output signal during the interval $0<t<T/10$, while the output from the other nine PSG units is zero for this interval. The PSG 2 unit delivers the entire output for the interval $T/10<t<T/5$, with the other units performing their functions in a similar manner. In this way, each PSG accurately and completely forms one segment of the resulting waveform.

Operation of the waveform generator should be relatively simple, if an oscilloscope is used. With the required contour for a given segment indicated on the screen by a mask or other contour, the corresponding PSG unit could be "knob-twisted" into the required shape with relatively little trouble after an initial learning period.

It is believed that n equal to 2 or 3 would give sufficient waveform accuracy for many applications. In this way the complexity of the network, and the amount of hardware needed in its construction would be reasonable. It seems that the one operational amplifier per A_0 unit in each of the PSG represents the main undesirable feature as regards equipment. It is recognized that these A_0 units may be conserved by increasing the degree of the polynomial through direct extension of the foregoing methods and thereby decreasing the number of PSG units needed to produce a given accuracy of waveform reproduction. This reduces the total number of operational amplifiers required, since only one is used per PSG unit.

Another possible form of the waveform approximation generator is that given in Fig. 18. The inverter network performs the function of producing $[-e_i(t)]$, given $[+e_i(t)]$ as an input signal. The operational amplifier method of producing this could be utilized, but this would increase the number of amplifiers required per waveform generator. The seriousness of the number of operational amplifiers in the final design would distinctly depend upon the capacity of the analog computer installation, and the magnitude of the problem being solved. It is possible that such refinements in the final design can be made that low-quality amplifiers may be built directly into the waveform approximation generator.

If only the first two terms of equation 11 were utilized in the PSGs, the waveform approximation generator would be the straight-line type mentioned earlier. Since A_0 and A_1 may be varied independently and made positive or negative in each segment of the generated wave, the resulting waveform approximation generator would possess complete flexibility. It would possess the same ease of experimental adjustment as was described for higher degree segment generators.

Conclusions

The delay-line Fourier synthesizer provides a relatively simple means whereby an irregular periodic waveform may be constructed from its Fourier series expansion. The accuracy with which the generated waveform adheres to the desired shape is limited by the number of terms of the Fourier series utilized, together with the accuracy with which the harmonics are generated.

The magnetic-core harmonic generator seems to offer particular promise as a tool for generating the harmonics needed in the delay-line Fourier synthesizer. The nonlinear-element harmonic generator offers a reasonably standard means of producing the harmonics, while the multiplier-type harmonic generator may prove to be of considerable value when improvements are made in the design of multipliers of relatively high-frequency range.

Function generation by waveform approximation provides an alternate method whereby an irregular periodic waveform may be produced. By this method, it is not necessary to know the Fourier series expansion of the function to be generated. For the generator types considered, it seems that an experimental cut-and-try method would produce the desired results in the minimum time.

A direct approach to straight-line waveform approximation using delay-line or short-time memory devices, when the input signal is a triangular wave, sawtooth wave, or square wave, seemed to produce systems which could handle only a restricted class of functions, when the duration of the nonzero part of the input signal was either the complete period of the desired wave or twice this value. It was found that a completely flexible straight-line approximation waveform generator could be designed using a periodic saw-tooth pulse input signal. The duration of the pulse is set equal to T/n, where T is the period of the wave to be generated, and n is the number of straight line segments utilized in generating the desired waveform.

The general form of the waveform approximation generator utilized the same periodic saw-tooth pulse input signal, but also utilized nonlinear networks so that each segment of the desired function can be approximated by a polynomial function of time of relatively high degree. The simplicity of the nonlinear network allows a high-degree polynomial to be generated with relatively inexpensive components.

It is hoped that the continued development of compact, efficient short-time memory devices will render some of the theoretical designs discussed both practical and economically justifiable. The high accuracy attainable in some of the designs may help to keep pace with the increasing accuracy reqnirements being asked of the analog computer industry.

References

1. ALTERNATING CURRENT MACHINES (book), A. F. Puchstein, T. C. Lloyd. John Wiley & Sons, Inc., New York, N. Y., second edition, 1942, p. 164.

2. MAGNETIC CIRCUITS AND TRANSFORMERS (book), Electrical Engineering Staff, MIT. *Ibid.*, 1947, p. 164.

3. TRANSVERSAL FILTERS, H. E. Kallman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 29, Mar. 1941, pp. 115–20.

4. A NONLINEAR COMPENSATING CONFIGURATION FOR SATURATING SERVOMECHANISMS, W. H. Surber, Jr. *Convention Record*, Institute of Radio Engineers, pt. 4, 1955, p. 23.

Watt-Hour Meter for Two Phases of a 4-Wire Y 3-Phase System

A. M. McQUARRIE
NONMEMBER AIEE

Synopsis: A newly developed watt-hour meter for measuring energy in two phases of a 4-wire Y 3-phase system is essentially a single-phase meter with L-R phase-shifting networks in the current circuits. This paper describes the meter and shows performance and operating characteristics.

MANY UTILITIES in the United States and Canada use in high-load-density areas a 120/208-volt secondary network system instead of a radial system. Network systems are selected primarily to serve commercial loads in high-density areas. Apartment dwellings in such areas served by the network can use 120-volt 2-wire service for small appliances and lighting, but the addition of major appliances such as room air conditioners, ranges, and water heaters, dictates higher voltages than 120 volts, and 3-wire service is required.

Three-wire 120/208-volt service, consisting of two phases of the 4-wire Y 3-phase network, is readily available, but such a system has required the use of a 2-stator polyphase type of watt-hour meter. Because the cost of such a meter is more than double that of a single-phase meter, it is natural that utilities with extensive 120/208-volt network distribution systems should view favorably any efforts by manufacturers to search for new and lower cost methods of metering these circuits.

The voltage-current phase relationship of 120/208-volt network loads prevents standard single-phase meters from registering such loads correctly. However, by the addition of simple phase-shifting networks, a 3-wire single-stator meter may be modified to meter this type of service. The lower-cost single-phase meter construction results in substantial cost savings over that of a 2-stator meter.

Description

The meter shown in Fig. 1 is based on this principle. It consists essentially of a double-current-coil single-phase meter with inductance-resistance phase-shifting networks connected in the current coil circuits. Fig. 2 shows the stator assembly.

The internal and external connections of such a meter with its associated networks are shown schematically in Fig. 3. In each phase, the current is divided, with a portion passing through the current coil of the meter and the balance shunted around the current coil. The potential coil is connected across the lines.

It is noted in Fig. 3, that both 208-volt line-to-line, and 120-volt line-to-neutral loads must be metered. Under unity power factor conditions, the current in the 208-volt load is in phase with the voltage applied to the potential coil of the meter. The currents in the 120-volt line-to-neutral loads, however, are out of phase with the potential coil voltage.

By choosing the proper values of inductances and resistances in the networks, the currents in the current coils of the meter are adjusted in magnitude and phase angle, so that for 120-volt unity power factor loading, the current fluxes have the following relationships, which are required for accurate metering: 1. They are in phase with each other. 2. They are 90 degrees out of phase with the potential flux. 3. They are a predetermined ratio to the line currents.

Fig. 4 shows the approximate relationships of the current coil currents to the line currents and to the voltage applied to the potential coil, for unity power factor line-to-neutral loads.

In the case of 208-volt line-to-line loads, the currents in the current coils are no longer in phase with the voltage applied to the potential coil. At unity power factor load, the current in one current coil leads the voltage by approximately 30 degrees, while the current in the other current coil lags the voltage, as is shown in Fig. 5.

Paper 58-145, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted March 27, 1957; made available for printing November 21, 1957.

A. M. McQUARRIE is with the General Electric Company, Somersworth, N. H.

Credit is due I. F. Kinnard for initial development work in 1927 on a meter based on the principle outlined in this paper.

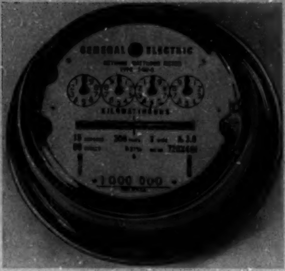


Fig. 1. Watt-hour meter for 120/208-volt network loads

Phase-Shifting Networks

In the development of this meter, the networks had to be made small and compact in order to keep the resistance low and the heating losses within acceptable limits. The minimum size of the reactors, however, is limited by the wire size necessary to carry the current without excess heating, the number of turns, the cross-sectional area of iron, and the air gap which is necessary to minimize the effect of the change of permeability of the iron with change of current. Furthermore, the reactors are proportioned to make use of the change of permeability with change of current for load compensation purposes.

The low impedance of the branches of the inductance-resistance networks (in the order of 0.003 ohm) requires that the contact resistance at the junction be kept very low and that it shall not change in service. For this reason, the connections are permanently silver-brazed in the factory. Adjustment of the individual load curves for phase balance and power factor is done at the factory. Field adjustment is limited to the usual full-load and light-load screws, which affect both phases together.

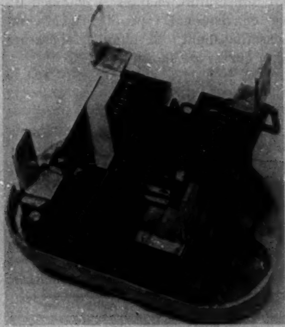


Fig. 2. Back view of stator assembly showing reactors and resistors

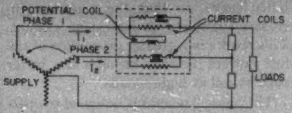


Fig. 3. Internal and external meter connections, showing phase shifting networks in current circuits

Balance between phases and lag adjustment are made by cutting between holes in the manganin resistance strips in the networks, thereby changing the values of resistance. Two such adjustable resistances are in each network, and the effect of changing these may be seen by reference to Fig. 6(B). The currents a_2 and b_2 are the currents in the two branches of the network in phase 2 with I_2 the resultant line current. The current a_2 through the current coil is maintained at approximately one half the line current I_2. The current b_2 is in phase with the voltage e_2 across the network, while a_2 lags by an angle ϕ, which is determined by the relative values of inductance and resistance in branch 2. The voltage e_2 across the network is shown only as a reference voltage to aid in understanding the relationships of the currents. It should be kept in mind that the line current is the forcing function in this network, and not the voltage e_2. By adjustment of r_2, the relative magnitudes of a_2 and b_2 are changed, as well as the angle ϕ between these currents. Adjustment of r_1 does not change the angle ϕ, but it does change angle θ, along with the relative magnitudes of the currents in the branches. It is seen that the two adjustments are not independent.

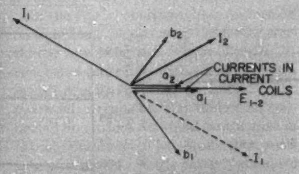


Fig. 4. Approximate relationship of currents in current coils to line currents and potential coil voltage for unity power factor line-to-neutral loads

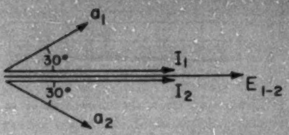


Fig. 5. Approximate relationship of currents in current coils to line currents and potential coil voltage for unity power factor line-to-line loads

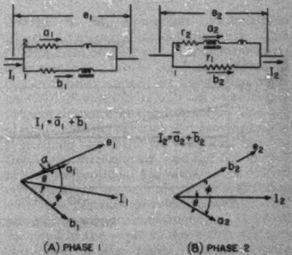
Fig. 6(A) shows a similar vector relationship for phase 1 in which $I_1 = a_1 + b_1$. In this case, due to the inductance of the current coil in branch 2, current a_1 is not in phase with the voltage e_1 across the network, but lags it by a small angle α.

The reactor required is not the same in each network. The inductance of the phase 1 reactor must be somewhat greater than that which was required in phase 2 for three reasons:

1. The reactor must compensate for the inductance of the current coil in the parallel branch.
2. A conventional type of lag plate is not used in the meter; hence, the useful potential flux lags the line voltage by an angle somewhat less than 90 degrees. Therefore the current in the current coil in phase 1 must lead the line current by a greater angle than would be required if the lag plate were retained.
3. The reactor must compensate for the lagging effect of the other network which is transformer-coupled through the current coil iron.

These factors operate in the opposite direction on the phase 2 circuit so as to reduce the amount of inductance required in the phase 2 reactor.

It is, of course, desirable to minimize the transformer coupling between the



(A) PHASE 1 (B) PHASE 2

Fig. 6. Relationship of currents in branches of networks to line currents

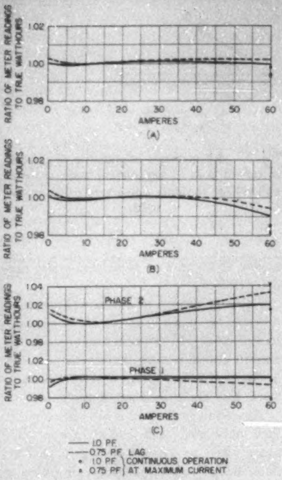


Fig. 7. I-62 meter load curves

A—Line-to-line loading
B—Line-to-neutral loading, phases energized
C—Line-to-neutral loading, individual phases

two networks in order to reduce the lagging effect and the heating losses. To help keep this coupling low, the conventional type of overload compensation is not used in this meter. Load compensation is obtained by making use of the change of permeability of the iron in the reactors with change of current.

The change in permeability of the iron at higher inductions has some effect on the calibration of the meter at higher loads, and this effect is used in the design of the reactors to compensate for the normal droop in the meter load curves. The phase 1 reactor is designed to operate only on the rising portion of the permeability curve, so that the reactance increases slightly with increasing current. This results in a greater proportion of the line current flowing through the meter current coil at higher loads.

The phase 2 reactor is designed to operate at higher flux densities at maximum line current. The lower iron permeability at maximum current causes a slight drop in the phase 2 reactance, and since the phase 2 reactor is in series with the phase 2 current coil, the meter load curve will tend to rise, as desired. In the case of the phase 2 reactor, however, the decrease in reactance with rising current is due mainly to increased flux leakage, and is indirectly due to the permeability change.

The low permeability of the iron in the reactors at low inductions tends to cause the meter to run fast on phase 2 and slow on phase 1 at light loads. The effect is kept within permissible limits by maintaining adequate air gaps in the magnetic circuits.

In addition to affecting the magnitudes of the currents in the current coil circuits, the change in impedance of the reactors with load introduces a small phase shift in the current coil currents. This affects the shape of the load curves at lagging power factors, as will be noted by referring to the load curves in Fig. 7(C). Low power factor loading on phase 2 will cause the meter to run fast at heavy loads; similar loads on phase 1 will cause the meter to run slow. With equal loading on both phases, these errors will tend to cancel, and the resultant low power factor load curve is essentially flat to maximum meter rating.

Temperature Effects

The resistances used in the networks are manganin strip, with a very low temperature coefficient. The copper in the reactance coils and in the current coils, however, has a temperature coefficient which will affect the magnitudes and phase angles of the currents in the two branches of each network, thereby introducing self-heat errors.

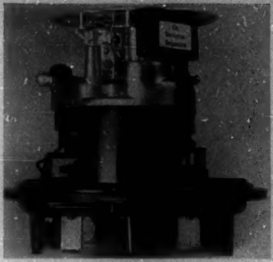


Fig. 9. Side view of meter, showing phase sequence indicator assembly

In the phase 1 network, Fig. 6(A), the copper is approximately the same in each branch. The self-heat error is therefore very small at unity power factor, but at low power factors an increase in temperature will cause an error in the slow direction, which amounts to about 2% at maximum current rating.

In phase 2, Fig. 6(B), copper is introduced into branch 1 of the network to compensate for the effect of the copper in branch 2. As in the other phase, the unity power factor self-heat error is thus made very small. Since the phase angle between the currents I_2 and a_2 tends to decrease with rise in temperature, there is an error at low power factors in the fast direction. This tends to balance out the low power factor temperature error in phase 1.

Temperature compensation for ambient temperature changes is obtained by the conventional type of magnet-compensating shunt. Resultant temperature error curves are shown in Fig. 8.

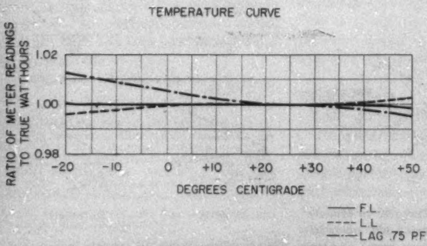


Fig. 8. I-62 meter temperature curves

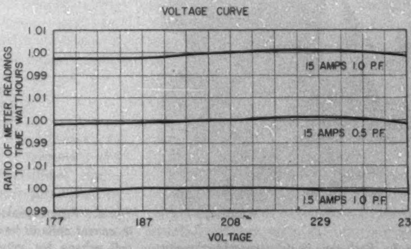


Fig. 10. I-62 meter voltage curves

Table I. Performance and Operating Characteristics

	Units	Type I-62 Single-Stator Network Watt-Hour Meter	Class 60 2-Stator Watt-Hour Meter
Rated voltage	volts	120/208	120/208
Nominal current rating	amperes	15	15
Maximum current	amperes	60	60
Frequency	cycles per second	60	60
Watt-hour constant	watt-hours per revolution	3.6	3.6
Speed at nominal load	rpm	$16\frac{2}{3}$	$16\frac{2}{3}$
Torque at nominal load	gram-millimeters	40	75
Potential circuit losses	watts	1.2	2.4
Starting watts	watts	10	18
Current ampere-turns (per stator)		60	90
Current circuit losses:			
At 15 amperes, approx	watts	1.25	0.68
At 60 amperes, approx	watts	20	10.9
External magnetic field error*	per cent	<±0.5	<±0.5
Weight of meter	pounds	$5\frac{3}{4}$	$8\frac{1}{4}$

* From *American Standard Code for Electrical Meters*, ASA C12-1941, conditions 1–3, section 4, paragraph 464.

External Magnetic Field Error

In this meter, each of the *L-R* networks comprises a closed loop, which includes a current coil of the meter. It might be expected therefore that the meter would be very sensitive to external magnetic fields.

However, these loops are arranged so that the fluxes induced in the meter current coils due to external fields are in opposite directions, and cancel each other. The net effect of external fields on meter accuracy is very small, as is shown in Table I.

Phase Sequence

Unlike conventional 2-stator meters, this meter will register line-to-neutral loads accurately only if the phase sequence is positive, as shown in Fig. 3. Therefore precautions are required on installation to ensure that the phase sequence is correct. In addition, it is necessary to assure that the phases are not reversed in service due to wiring changes in other parts of the system.

To take care of this problem, one model of the meter is being developed with a phase rotation indicator incorporated in the form of a neon lamp which glows if the phase sequence is in the wrong direction. The phase rotation indicator assembly is shown in Fig. 9.

The phase rotation indicator consists of two resistors and a capacitor connected across the 208-volt supply, with the neon lamp connected from the junction of the two resistors to ground. The values of resistances and capacitance are so chosen that, with incorrect phase rotation, 120 volts are applied across the neon lamp, causing it to light. With proper phase rotation the voltage across the lamp is approximately zero.

Thus, the meter installer is able to determine readily, at the time of installation, whether the phase rotation is correct. If by any chance the phase sequence should become reversed, due to wiring changes in other parts of the service, the neon lamp will glow and warn of the condition.

Performance and Operating Characteristics

The performance and operating data of the single stator network meter are compared with that of a typical 2-stator network meter of the same rating in Table I. Characteristic curves of the single stator meter are given in Figs. 7, 8, and 10.

Calibration and Test Procedure

The field test procedure for the single stator network meter is very similar to that for a single-phase 3-wire meter, with the two current circuits connected in series. As mentioned, the only adjustments are those of full load and light load, which are made in the same manner as with a single-phase meter. Thus, the routine field testing of the meter is much faster and more economical than that of a 2-stator meter.

In the factory, calibration is performed on individual phases with the meter connected to a balanced voltage 3-phase Y-connected generator, with all three phases loaded equally from line to neutral. The potential circuit of the meter is connected line to line, as would be the case in actual operation. The standard meter is a conventional single-phase standard type, and is inserted in series with the particular current circuit being adjusted. The potential circuit of the standard is connected from phase to neutral. For lag calibration, 0.5 power factor is obtained by cross phasing.

With the voltage on the standard being of a different magnitude and at a 30-degree phase angle from the voltage applied to the meter under test, the importance of precise control of these supply voltages can be understood. However, the separate 3-phase Y-connected generators supplying each test table take care of this problem automatically, with no attention from the operator.

Although no independent adjustment of the individual phases is possible in the field, a spot check of the calibration of each phase by the utilities is, of course, desirable. Such checks may readily be made with the use of single-phase meter standards, provided the proper voltage magnitude and phase relationships are maintained between the meter being tested and the standard.

Summary

This paper shows how the addition of phase-shifting networks to the current circuits of a 3-wire single-stator meter makes it applicable for measurement of power in two phases of a 4-wire Y 3-phase system.

The meter is proposed for domestic network system loads, where voltages are balanced, and power factor is normally higher than 0.75. Such loads include window-type air conditioners and major electric appliances, as well as small loads such as clocks and night lights. The smaller loads will be applied on individual phases, whereas the heavier loads will normally be on both phases. The performance is shown to compare very favorably with more costly 2-stator meters.

◆

Discussion

George B. M. Robertson (Philadelphia Electric Company, Philadelphia, Pa.): The development of the 3-wire single-stator watt-hour meter for use in the measurement of line-to-neutral as well as line-to-line loads on 120/208-volt 3-wire services represents a distinct advance in the industry. The use of phase-shifting networks for accomplishing this result represents an ingenious innovation in the metering art, and the elimination of the conventional lag plate and overload compensation by combining this function with the reactors provides desirable simplification in meter construction. The data included in the paper indicate that the performance of the meter compares quite favorably with that of the 2-stator meters that have been necessary in the past for correct measurements on this type of service.

Not included in the paper are data on the effects of variation in frequency. This may have been omitted because the stability of the large interconnected systems of today is such as to render this feature of negligible importance. However, the effects of frequency variation serve as a valuable index to meter performance on circuits having harmonics. I would therefore ask Mr. McQuarrie whether he has any test data on the effects of frequency variation.

From the point of view of practical operation it may be well to stress that while the meter will measure correctly line-to-neutral and line-to-line loads on two phases of 120/208-volt 3-phase 4-wire circuits, it cannot be used for other types of service which are at times metered with the conventional 2-stator network meter. Thus the new meter cannot be used on 3-phase 3-wire, 2-phase 3-wire, or single-phase 3-wire circuits. However, for the purpose for which it is intended, the use of the new meter will represent a distinct economy over the 2-stator meter for those utilities that have extended 4-wire Y 3-phase systems.

C. V. Morey (Consolidated Edison Company of New York, Inc., New York, N. Y.): The development and production of a commercially practicable single-stator watt-hour meter for measurement of 3-wire service from a 3-phase 4-wire 120/208-volt distribution system is a revolutionary step forward in watt-hour metering and has a decided significance for utility companies from the standpoint of economics. The General Electric Company is to be commended for its accomplishment, and the author for his clarity in presenting the technical details of construction and operation.

The economic factor is particularly important in the case of the Consolidated Edison Company of New York, since we have over 2,000,000 customers supplied from a 3-phase 4-wire system.

In the early days of 3-phase, 4-wire networks, the only meter available for 120/208-volt (so-called open-Y) service, was the bulky, cumbersome, and rather expensive 2-stator polyphase meter normally used for 3-phase delta or 2-phase power measurement. Simplified and more compact revisions of this 2-stator meter were produced later on and with the continuous advancement of the metering art, 2-stator meters are presently available in mounting frames and housings not much larger than those for single-stator meters. However, meters of the 2-stator construction are naturally more costly to build than single-stator meters, and this differential is further increased because of much lower volume of production. Such meters, for 3-wire 120/208-volt application, have generally sold for about $2^{1}/_{2}$ times the price of single-stator meters for 2-wire 120-volt or 3-wire 120/240-volt application. Because of this difference in price, new residential customers on the Consolidated Edison Company system were, up to 3 or 4 years ago, practically all served at 120 volts, 2-wire, so that less costly single-stator 2-wire meters could be utilized.

With increased use of heavy appliances and room air conditioners, however, 2-wire service came to be considered inadequate in our territory for residential loads. This inadequacy applied both to new installations and to existing cases where customers were rewiring in order to provide large risers for apartment house tenants. The potential requirement for new meters for 3-wire 120/208-volt installations thus became extremely large, the rate of use of such meters now being in excess of 50,000 per year. As a consequence, the meter manufacturers were urged to develop a more economical meter which would be suitable for this application, and the type *I-62* meter is the result of the successful effort made by the General Electric Company in this direction.

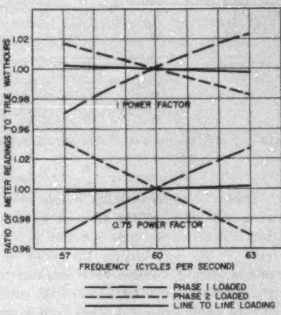


Fig. 11. I-62 meter frequency curves

The only practical difficulty with this new meter is the necessity for wiring the two phase conductors to it in a particular phase sequence and ensuring that such phase sequence is maintained as long as the meter remains in service. As stated in the author's paper, phase sequence must be correct or 120-volt 2-wire loads will not be registered properly. This difficulty has been overcome in the case of Consolidated Edison Company by making use of the color-coded wiring required to be installed by contractors. The colors are white, black, and red for 3-wire service, and white, black, red, and blue for 4-wire service; white being the neutral conductor in both cases and the other colors the phase conductors. Where the supply to a single or multifamily building is 3-wire, the color combination of the phase conductors will always be red-black. Where the supply to the building is 3-phase, 4-wire, three different color combinations (red-black, black-blue, blue-red) may be derived to supply 3-wire service to different customers. Wiring prints were issued to all contractors instructing them as to how each color combination should be connected to meter-mounting terminals. When the Company connects the colored phase conductors to the distribution system at the service point, the connection is made so that black leads red where the service is 3-wire, and so that black leads red leads blue where the service is 3-phase 4-wire. To aid contractors and also Company inspectors, the Company developed a new form of phase sequence indicator, about as large as a good-sized fountain pen, and had it made in quantity. This device is used to make a final check for phase sequence at the meter mounting before the meter is installed.

The addition of the neon lamp indicator referred to in the paper, the circuit of which is constructed along the same lines as the phase sequence indicator just mentioned, will afford a convenient means for meter indexers to note and report any case where the phase sequence has been accidentally or intentionally reversed.

The new meter, besides costing less, is smaller, lighter, and easier to handle than a conventional 2-stator meter and the fact that adjustments for power factor and equality of circuits are preset by the manufacturer makes it about equivalent to a 3-wire 120/240-volt single-stator meter for simplicity of test and calibration.

A. M. McQuarrie: Mr. Robertson's question with respect to the effects of variation in frequency can best be answered by reference to the frequency error curves shown in Fig. 11.

As would be expected, because of the inductances in the phase-shifting networks, the individual phases of the meter are somewhat frequency sensitive. However, since the frequency errors on the individual phases are in opposite directions, these errors cancel out when both phases are loaded. The result is a practically flat frequency error curve when the phase loads are balanced or when the load is 208 volts line to line.

Mr. Morey has shown how the correct phase rotation at the meter may be ensured during installation by means of color coded wiring. Such a system should be applicable by any utility considering the use of single-stator network meters, thus minimizing the application difficulties.

A Stabilized D-C Differential Transistor Amplifier

LOUIS DEPIAN
ASSOCIATE MEMBER AIEE

RICHARD E. SMITH
ASSOCIATE MEMBER AIEE

TRANSISTORS present many advantages over vacuum tubes. Though they may be substituted for tubes in amplifiers, oscillators, pulse generators, pulse counting circuits, etc., they only amount to one hundredth of the weight of tubes, one thousandth of the volume, they consume one tenth of the power, and their average life is shown to be longer.

In certain fields, however, their application has been limited by their inherent sensitivity to temperature. This is particularly true in d-c and switching amplifiers, where changes in the amplifier operating point may cause outputs which are indistinguishable from those caused by actual input signals.

Temperature Variations of Transistor Parameters

An elementary, transistor d-c amplifier, as shown in Fig. 1, may be characterized by three parameters α_{cb}, I_0', and r_d. Each of these parameters depends, in general, on temperature.

The amplification factor for the common emitter circuit α_{cb} can be shown to be nearly independent of temperature.[1] Fig. 2 shows an experimental curve for the variation of α_{cb} as a function of temperature. While α_{cb} is nearly independent of temperature, r_d does show a small variation as the ambient temperature is changed (Fig. 3).

These variations are insignificant compared to the dependence of I_0' on temperature. I_0' is the collector current when the base current is zero and the collector-to-emitter voltage is of the order of magnitude of a few volts. It is related to I_{co}' (collector current when emitter current is zero and when collector-to-base voltage is of the order of magnitude of a few volts) by

$$I_0'=I_{c0}'(1+\alpha_{cb})$$

It can be shown[2] that I_{co}', and consequently I_0', will increase at the rate of about 10% per degree centigrade at room temperature. Fig. 4 shows experimental curves for the variation of I_0' as a function of temperature. It is evident that the variation of I_0' with temperature is much greater than that of α_{cb} or r_d, and it is thus the primary cause of temperature dependence of the operating point. Any practical d-c or switching amplifier must compensate for this variation.

If, in the amplifier of Fig. 1, I_0' is increased by a factor of b, the operating point will drift, and the collector current will increase by a factor of a

$$\alpha=\frac{1+\lambda b}{1+\lambda} \tag{1}$$

where

$$\lambda=\frac{r_d I_0'}{E_{ce}}\,\frac{1}{1+\dfrac{\alpha_{cb} r_d}{R_f}} \tag{2}$$

This variation in i_c is shown in Fig. 5. Changes in the collector current with variation in α_{cb} and r_d may be calculated in a similar manner.

Compensating Circuits

There have been several circuits devised for temperature stabilization of d-c transistor amplifiers. One of the more common methods is use of direct voltage and current feedback.[1] The voltage feedback methods are limited in their use to cases where the d-c resistance in the collector circuit is high. On the other hand, the current feedback method has a greater range of design freedom but is limited to cases where the d-c resistance in the collector circuit is low. In general it may be said that feedback achieves temperature stabilization at the expense of current gain.

Feedback in individual stages of amplification may be used to compensate partially for transistor-parameter variations. Two stages may then be coupled together so that the parameter variations in one stage will modify the feedback in the other stage and temperature-stabilize the amplifier.[3] This method, however, requires the use of transistors with a high-current amplification factor and low leakage current. Furthermore, none of these circuits provides zero output with zero input unless special circuitry has been added.

Other methods of stabilization make use of nonlinear temperature characteristics of other devices for compensation.[4-6] However, these have the disadvantage that each transistor requires a matched element for temperature compensation.

The Differential Amplifier

It is shown in the foregoing that the main variation in transistor parameters occurs in I_0. The method described here eliminates this variation by employing two transistors in the amplifier shown in Fig. 6. This circuit accomplishes the compensation by subtracting the voltages due to I_{01}' and I_{02}' across R_{L1} and R_{L2}, respectively. On the other hand, the effect of an actual input signal is not cancelled and will appear as an output. The output voltage V_0 is usually given by

$$v_0=i_s\left(\frac{\alpha_{cb_1}R_{L_1}}{1+R_{L_1}/r_{d_1}}+\frac{\alpha_{cb_2}R_{L_2}}{1+R_{L_2}/r_{d_2}}\right)+ E_{ce}\left(\frac{1-\alpha_{cb_1}R_{L_1}/R_{f_1}}{1+R_{L_1}/r_{d_1}}+\frac{1-\alpha_{cb_2}R_{L_2}/R_{f_2}}{1+R_{L_2}/r_{d_2}}\right)+ I_{0_1}'\frac{R_{L_1}}{1+R_{L_1}/r_{d_1}}-I_{0_2}'\frac{R_{L_2}}{1+R_{L_2}/r_{d_2}} \tag{3}$$

For compensated operation only the first term should exist and the other be eliminated. This is readily achieved if

$$\frac{R_{L_2}(1+R_{L_1}/r_{d_1})}{R_{L_1}(1+R_{L_2}/r_{d_2})}=\frac{I_{0_1}'}{I_{0_2}'} \tag{4}$$

and

$$\frac{1-\alpha_{cb_2}R_{L_2}/R_{f_2}}{1+R_{L_2}/r_{d_2}}=\frac{1-\alpha_{cb_1}R_{L_1}/R_{f_1}}{1+R_{L_1}/r_{d_1}} \tag{5}$$

If $R_{L_1}/r_{d_1}<<1$ and $R_{L_2}/r_{d_2}<<1$, the above expression can be simplified

$$\frac{R_{L_2}}{R_{L_1}}=\frac{I_{01}'}{I_{02}'} \tag{4A}$$

and

$$\frac{\alpha_{cb_2}R_{L_2}}{\alpha_{cb_1}R_{L_1}}=\frac{R_{f_2}}{R_{f_1}} \tag{5A}$$

effectively eliminating the effects of r_d variations, since the output voltage is then

$$V_0=i_s(\alpha_{cb_1}R_{L_1}+\alpha_{cb_2}R_{L_2}) \tag{6}$$

Notice that both transistors contribute in current amplification. This circuit eliminates the effect of r_d variations only for small values of R_L (of the order of magnitude of $1K$). If higher values of R_L are desirable, the circuit may be modified, as shown in Fig. 7. Here the output voltage is

---

Paper 58-192, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 23, 1956; made available for printing December 9, 1957.

LOUIS DEPIAN is with George Washington University, Washington, D. C., and RICHARD E. SMITH is with Carnegie Institute of Technology, Pittsburgh, Pa.

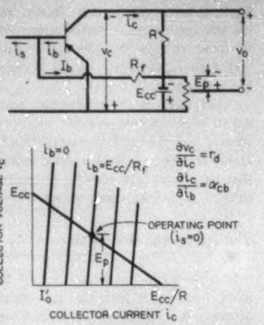


Fig. 1. Conventional common-emitter PNP junction transistor amplifier and characteristic curves

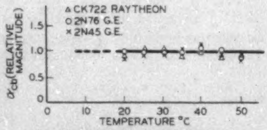


Fig. 2. Current-amplification factor α_{cb} (common emitter) as function of temperature

$$V_0=i_s\left(\frac{\alpha_{cb_1}R_{L_1}}{1+\beta_{01}}+\frac{\alpha_{cb_2}R_{L_2}}{1+\beta_{02}}\right)+E_{cc}\left(\frac{1}{1+\beta_{01}}-\frac{1}{1+\beta_{02}}\right)+I_{01}'\frac{R_{L_1}}{1+\beta_{01}}-I_{02}'\frac{R_{L_2}}{1+\beta_{02}} \quad (7)$$

where

$$\beta_{01}=R_{L_1}\left(\frac{1}{R_{f_1}}(1+\alpha_{cb_1})+\frac{1}{r_{d_1}}\right)$$

$$\beta_{02}=R_{L_2}\left(\frac{1}{R_{f_2}}(1+\alpha_{cb2})+\frac{1}{r_{d_2}}\right) \quad (8)$$

For compensation

$$\beta_{01}=\beta_{02} \quad (9)$$

and

$$\frac{R_{L_2}}{R_{L_1}}=\frac{I_{01}'}{I_{02}'} \quad (10)$$

From equation 8 it can be seen that the elimination of the r_d variation may be effected through R_{f_1} and R_{f_2} rather than through R_{L_1} and R_{L_2}. If

$$\frac{\alpha_{cb}}{R_f}>>\frac{1}{r_d} \quad (11)$$

then equation 9 is simplified to

$$\frac{\alpha_{cb_1}}{\alpha_{cb_2}}\frac{R_{f_2}}{R_{f_1}}=\frac{R_{L_2}}{R_{L_1}} \quad (9A)$$

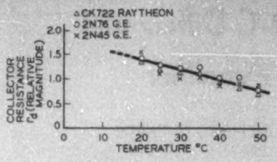


Fig. 3. Collector resistance r_d (common emitter) as function of temperature

The output voltage is then

$$V_0=i_s\left(\frac{\alpha_{cb_1}R_{L_1}}{1+\alpha_{cb_1}R_{L_1}/R_{f_1}}+\frac{\alpha_{cb_2}R_{L_2}}{1+\alpha_{cb_2}R_{L_2}/R_{f_2}}\right) \quad (12)$$

Notice that the output voltage here is lower than that of equation 6.

The amplifiers of Figs. 6 and 7 require different values of R_L and R_f for each transistor in order to achieve stabilization.

It is possible to obtain compensation with the same value for R_L for each transistor by the insertion of a small resistance, as shown in Fig. 8.

The compensation criteria here are

$$1+\frac{R_0}{r_{d_1}}=\gamma=\frac{I_{01}'}{I_{02}'} \quad (13)$$

and

$$\frac{\alpha_{cb_1}}{\gamma R_{f_1}}=\frac{\alpha_{cb_2}}{R_{f_2}} \quad (14)$$

provided that

$$\frac{\alpha_{cb}}{R_f}>>\frac{1}{r_d} \quad (15)$$

Then

$$V_0=i_s\left(\frac{\alpha_{cb_1}R_L}{1+\alpha_{cb_1}R_L/R_{f_1}}+\frac{\alpha_{cb_2}R_L}{1+\alpha_{cb_2}R_L/R_{f_2}}\right) \quad (16)$$

Equation 13 shows that if the variation of r_d with temperature is pronounced, the amplifier of Fig. 7 should be used rather than that of Fig. 8.

The amplifiers described are capable of supplying power to a load resistance connected across their output terminals. In all the circuits described the conditions for temperature stabilization are the same as those obtained for the unloaded cases. The only modification is a factor required for the output voltage. For example, if a load resistance R_c is connected across the output terminals of the amplifier shown in Fig. 6, the output voltage expression in equations 8 becomes

$$V_0'=\frac{1}{\alpha}V_0 \quad (17)$$

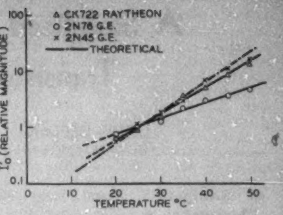


Fig. 4. I_0' as function of temperature

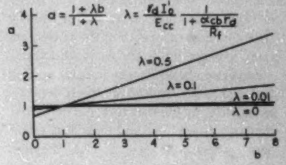


Fig. 5. Variation of collector quiescent current versus variation of I_0'

where

$$\alpha=1+\frac{R_{L_1}}{R_c(1+R_{L_1}/r_{d_1})}+\frac{R_{L_2}}{R_c(1+R_{L_2}/r_{d_2})}$$

If the specified conditions for equations 4(A) and 5(A) to hold are satisfied, then the output voltage V_0' is simplified to

$$V_0'=i_s\frac{R_c}{R_c+R_{L_1}+R_{L_2}}(\alpha_{cb_1}R_{L_1}+\alpha_{cb_2}R_{L_2}) \quad (18)$$

If a load R_c is applied to the amplifier shown in Fig. 8, the output voltage becomes

$$V_0'=\frac{R_c i_s}{R_c+\frac{R_L}{\gamma+\alpha_{cb_1}R_L/R_{f_1}}+\frac{R_L}{1+\alpha_{cb_2}R_L/R_{f_2}}}\left(\frac{\alpha_{cb_1}R_L}{\gamma+\alpha_{cb_1}R_L/R_{f_1}}+\frac{\alpha_{cb_2}R_L}{1+\alpha_{cb_2}R_L/R_{f_2}}\right) \quad (19)$$

One may obtain the output voltage in the loaded case for the amplifier shown in Fig. 7 from equation 19 by setting $\gamma=1$ and putting in the appropriate R_{L_1}'s and R_{L_2}'s where they belong. Thus, by the appropriate use of equations 17, 18, and 19, it is possible to obtain the current gain for any single-stage differential amplifier with any load, and, by the proper use of the conditions mentioned in this paper, to build a temperature-stable d-c transistorized amplifier. This also indicates that it is possible to cascade these amplifiers without any difficulty.

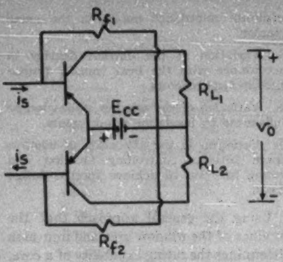


Fig. 6. Differential amplifier with fixed bias

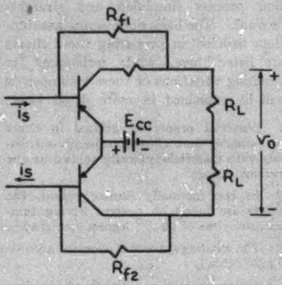


Fig. 7. Differential amplifier with self-bias

Design of a Single-Stage Differential Amplifier

To design a differential amplifier it is necessary to know the current gain and the load resistance to be used. By making some rough approximations and using equations 18 and 19 it is possible to arrive at some rough figures for α_{cb}, and thus choose a transistor. An examination of these two expressions makes clear that this is the only parameter that affects the gain of the difference amplifier. When a transistor has been selected, the parameters for the two transistors must be determined. It is actually only necessary to determine experimentally the I_0's for the transistors by measuring the collector current when the base current is zero and the collector-to-emitter voltage is of the order of a few volts. The other parameters can be calculated from manufacturers' curves. Then from equations 4(A), 10, or 13 the ratio of the R_L's is obtained, or the value of R_0, as the case may be. Of course, it is necessary to check whether the conditions for the application of each of these equations are met by the rough design already outlined, but this is easy to do once the transistor is chosen and the manufacturer's curves are obtained.

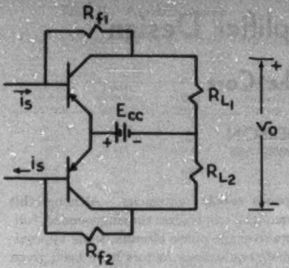


Fig. 8. Modification of differential amplifier with self-bias

With the determination of the R_L's or R_0 one can then refine the original, rough design by using the methods already available for vacuum-tube amplifiers and the expressions given in the foregoing. When the amplifier is completed, the second adjustment, indicated by equations 5(A), 9(A) and 14, can be made experimentally by varying one of the bias resistors so that the output voltage is zero when the input signal is zero.

If a double power supply is available, the differential amplifier suggested by Hurley[1] can be used.

Experimental Results

A differential amplifier was constructed by the use of the procedure outlined. The current gain was measured as a function of ambient temperature. The results of this measurement are shown in Fig. 9.

For a demonstration of the effectiveness of this method, one transistor was removed, and the output current was experimentally determined for the single-transistor amplifier, and then for the differential amplifier. The results are plotted in Fig. 10.

Conclusions

This paper shows that reliable temperature-stable, d-c transistor amplifiers can be built with differential amplifiers. It is, of course, possible that other circuits be used as well, but the differential amplifier does not require any highly special measurements to be set up, nor does it have severe limitations concerning loads. The procedure allows the designer much freedom, and requires no tedious calculations. Each transistor that is used contributes to the gain of the entire amplifier, and no device is employed as a temperature compensator alone. Stages may be cascaded as easily as those of a single-stage d-c amplifier, and these interactions can be predicted from the expressions indicated.

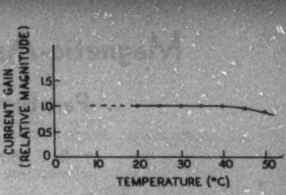


Fig. 9. Current gain as function of temperature for differential amplifier

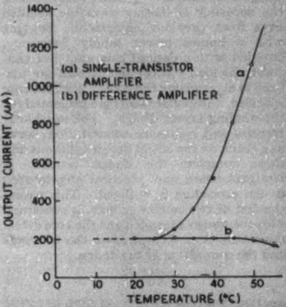


Fig. 10. Temperature stability of single-transistor amplifier and of differential amplifier, compared

References

1. TRANSISTOR ELECTRONICS (book), **A. W. Lo, R. O. Endres.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1955.

2. P-N JUNCTION TRANSISTORS, **W. Shockley, M. Sparks, G. K. Teal.** *Physical Review*, New York, N. Y., vol. 83, July 1951, p. 158.

3. TRANSISTORIZED D. C. AMPLIFIERS, **J. W. Stanton.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-3, no. 1, Mar. 1956, p. 65.

4. TEMPERATURE STABILIZATION OF TRANSISTOR AMPLIFIERS, **L. M. Vallese.** *Research Report R-470-56, PIB-400*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1956.

5. D-C AMPLIFIER EMPLOYING JUNCTION-TYPE TRANSISTORS, **Edward Keonjian.** *Electrical Engineering*, vol. 72, Nov. 1953, pp. 961–64.

6. TEMPERATURE-COMPENSATED D. C. TRANSISTOR AMPLIFIER, **Edward Keonjian.** *Proceedings*, Institute of Radio Engineers, vol. 42, Apr. 1954, pp. 661–77.

7. DIRECT-COUPLED TRANSISTOR AMPLIFIERS, **R. B. Hurley.** *Electronic Equipment*, London, England, vol. 3, Apr. 1955, pp. 26–28.

8. THE EMITTER-COUPLED DIFFERENTIAL AMPLIFIER, **D. W. Slaughter.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. CT-3, no. 1, Mar. 1956, p. 51.

Magnetic-Amplifier Design

Part I—The Core

R. E. ANDERSON
ASSOCIATE MEMBER AIEE

Synopsis: A practical approach is developed for the design of the toroidal magnetic amplifier, based on synthesizing the design from design charts, and thus avoiding the necessity of having available detailed data from previous equipment. Starting with a known power supply frequency, together with a desired gain and time constant, the load voltage and current are utilized in determining the core. With the core specified, additional charts are presented for the winding layout design. The usual heat transfer and resistance computations serve as a check on the initial design synthesis and these complete the design process. A straightforward and practical step-by-step design procedure is outlined. This paper, the first of two on this subject, is concerned with the design approach and the core determination. Part II will cover the windings and the completion of the design.

THE EXCELLENT SUITABILITY and fine performance of the magnetic amplifier in applications requiring ruggedness and reliability have led to increasing use of this device as a control and instrumentation component. From this situation arises a need for information and methods to be used in the design of the magnetic amplifier as a system component. A number of articles[1-8] have discussed specific problems in the design area, while other authors[9-10] have discussed some of the general aspects of design. In view of this background, it is the purpose here to present an organized collection of data which may be used in a practical step-by-step manner to design the toroidal magnetic amplifier, thus bringing together the information necessary for undertaking the design process. These data will cover magnetic amplifiers of all sizes for operation with various power supply frequencies. Although this material emphasizes the single-ended, full-wave single-phase circuits, some approximate equivalence factors have been given for various other circuits. The scope of this work will be limited to the self-saturating type of magnetic amplifier. The material is directed toward designing the circuit and reactor component while leaving information on core materials, wire, rectifiers, and heat transfer to commercially available information and the published literature in these fields of technology.

Design Technique

A group of charts has been prepared to permit the design of a toroidal amplistat without reference to previous designs. Perhaps it would be in order to discuss the design philosophy on which these charts are based and the principles which were employed in the preparation of the charts. Of course, it is well known that the size of an electromagnetic device depends upon its power rating or its volt-ampere rating. The exact manner in which the rating influences the size of a magnetic amplifier might be explained a little more clearly as follows: In a given reactor, to be used as a component in the magnetic amplifier circuit, certainly it is evident that the voltage rating is directly proportional to the cross-sectional area of the iron multiplied by the number of turns; while the current-carrying capacity is directly proportional to the area of the window, or of the copper, used in the conductors divided by the number of turns. Thus, it may be seen that the product of the window area times the iron area will be proportional to the volt-ampere rating of the reactor. (For more details see Appendix I.) In practice, naturally, this relationship is tempered by the fact that there are certain parameters whose magnitude varies with the size of the reactor. Some of the more important of these factors may be listed as follows:

1. Variation of the stacking factor with different thicknesses of core material.
2. Variation of the current density to maintain impedance match in the small reactor.
3. Variation of the current density in accordance with the heat transfer limitations in large reactors.
4. Variation of the window utilization as influenced by insulation requirements.
5. Variation in the allocation of window space between controlling function and gating function to achieve specific design goals.

Using the general approach that the product of the window area and iron area determines the rating capability of a core, together with laboratory meaurement of the factors just listed, design charts have been prepared to make the design calculation process simplified and straightforward. The basic engineering assumptions utilized in preparing these charts are listed here, while techniques for handling variations of these assumptions will be explained in more detail later:

1. Material properties utilized in chart preparation were chosen to be commensurate with materials typically applied at the various ratings.
2. In the thermally limited region, the charts are based on designs having temperature rises of 55 C (degrees centigrade).
3. The winding space was allocated according to Table I.
4. Approximately one eighth of the line voltage is assumed lost to after-saturation inductance. (Smaller drops may be achieved under favorable conditions.)
5. The chart is based on a typical design insulated for a 1,500-volt hi-pot (high-potential) test.
6. The copper area is assumed 20% allocated to all control functions, leaving 80% for the gating function.

To give the reader an understanding of the organization of this paper, the design procedure will now be summarized briefly before proceeding to the specific details.

From the rating of the magnetic amplifier, a core determination chart is used to determine the product of the core window

(*Continued on page 164*)

Table I. Allocation of Window Space

Item	Space Allocated, Per Cent of Total Window
Hole after winding	40.
Copper	23.
Wire insulation	16.5
Coil insulation	10.
Spacing*	7.5
Winding, randomness	3.
	100.0

* Space lost because wire has a circular cross section, while the available space is approximately trapesoidal in cross section.

Paper 57-922, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted April 20, 1957; made available for printing October 18, 1957.

R. E. ANDERSON is with the General Electric Company, Schenectady, N. Y.

The author wishes to thank Dr. H. F. Storm of the General Electric Company for his provident counseling and helpful guidance which were so essential to the preparation of this material. Acknowledgment is also due many of the author's colleagues in the General Electric Company for their co-operation in making available detailed design data on those magnetic amplifiers which had exhibited successful production performance.

Table II. Available Cores

Sizes and Catalog Numbers of Tape-Wound Toroidal Cores

					CORE SIZE			CASE SIZE			
Item No.	Window Area x Iron Area (inches)^4	Iron Area (inches)^2	Window Area (inches)^2	Mean Length of Magnetic Circuit cm.	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	Catalog Number*
1	.00029625	.003875	.076453	3.23809	.437	.375	.125	.437	.312	.198	AE-11
2	.00035343	.005	.070686	3.31155	.455	.375	.125	.455	.300	.215	MM-CA
3	.00042912	.015625	.027464	2.492375	.375	.250	.250	.375	.187	.328	AE-3
4	.00043488	.0039375	.110446	3.73657	.500	.437	.125	.500	.375	.198	AE-5
5	.0004753	.003125	.1521	4.18	.550	.500	.123	.610	.440	.195	MI-50158
6	.00055225	.005	.11045	3.949928	.535	.455	.125	.535	.375	.215	MM-BA
7	.00058117	.003875	.14998	4.235043	.562	.500	.125	.562	.437	.198	AE-12
8	.0006088	.007813	.07793	3.49	.500	.375	.125	.560	.315	.195	MI-50153
9	.00070686	.010	.070686	3.31155	.455	.375	.250	.455	.300	.340	MM-CC
10	.0009086	.007813	.1163	3.99	.563	.438	.125	.622	.377	.195	MI-50154
11	.00096123	.003875	.24806	5.23199	.687	.625	.125	.687	.562	.198	AE-13
12	.0011045	.010	.11045	3.949928	.535	.455	.250	.535	.375	.340	MM-BC
13	.0011534	.005875	.19634	4.85714	.656	.562	.125	.656	.500	.198	AE-6
14-H	.001188	.007813	.1521	4.49	.625	.500	.125	.685	.440	.195	MI-50056
15	.0012448	.005	.24895	5.450103	.723	.643	.125	.723	.563	.215	MM-AA
16	.001325	.0098	.13526	4.42	.633	.475	.125	.710	.415	.195	MM-1
17	.001817	.01563	.1163	3.99	.563	.438	.250	.622	.377	.330	MI-50155
18	.00189	.007812	.2419	5.49	.750	.625	.125	.810	.555	.195	MI-50057
19-L	.002265	.0156	.14522	4.98	.750	.500	.125	.820	.430	.195	AE-5340, MM-2, MI-5000
20	.002295	.01169	.1963	5.25	.750	.563	.125	.810	.500	.195	MI-50138
21	.0023434	.015625	.14998	4.486275	.625	.500	.250	.625	.437	.328	AE-14
22	.0024895	.010	.24895	5.450103	.723	.643	.250	.723	.563	.340	MM-AC
23	.0025954	.005875	.44178	7.6486	.906	.812	.125	.906	.750	.198	AE-7
24	.002832	.0078	.36316	6.50	.875	.750	.125	.945	.680	.195	MI-50143
25	.003427	.0132	.25967	5.89	.844	.633	.125	.915	.575	.195	MM-4
26	.00378	.01563	.2419	5.99	.875	.625	.125	.945	.555	.195	MI-50063
27-M	.004122	.0156	.26420	6.18	.900	.650	.125	.970	.580	.195	AE-5515, MM-5, MI-50002
28	.004545	.0313	.14522	4.99	.750	.500	.250	.820	.430	.320	AE-6592, MI-50176, RZ-9
29	.005298	.0078	.67929	8.49	1.125	1.000	.125	1.195	.930	.195	MI-50086
30	.005665	.0156	.36316	6.99	1.000	.750	.125	1.070	.680	.200	MI-50134
31	.00756	.03125	.2419	5.99	.875	.625	.250	.945	.555	.330	MI-50033, RZ-3
32	.00772	.031	.24894	5.490	.813	.563	.250				RL-33
33	.008527	.03525	.2419	6.49	1.000	.625	.188	1.070	.555	.268	MI-50076
34	.008534	.0235	.36316	6.74	.938	.750	.250	1.010	.680	.330	MI-50003
35	.00923	.047	.19634	5.004	.750	.50	.375				WC-1001
36	.01026	.0156	.65755	8.99	1.250	1.0	.125	1.340	.915	.195	MM-9, AE-4168, MI-50011
37-L	.0108	.0352	.307	6.49	1.000	.6250	.1875				AIEE
38	.01094	.0469	.23328	6.48	1.000	.625	.250	1.080	.545	.320	AE-5651, MI-50007, MM-3
39-M	.01137	.0313	.36316	6.99	1.000	.750	.250	1.070	.680	.330	MI-50061
40	.01170	.047	.24894	5.989	.938	.563	.250				RL-36
41	.01197	.0264	.45364	7.85	1.125	.844	.188	1.215	.760	.262	MM-8
42	.01208	.0353	.34211	7.48	1.125	.750	.188	1.215	.660	.262	AE-5504, MM-7, MI-50106
43	.01237	.063	.19634	5.98475	1	.50	.25				RH-16
44	.01369	.055	.24894	6.236	1	.563	.25				RL-9
45	.01512	.0625	.24192	6.97	1.125	.625	.250	1.195	.555	.320	MI-50167
46	.015531	.015625	.99401	10.467975	1.375	1.250	.250	1.375	1.125	.390	AE-8
47-L	.01555	.0352	.4418	7.4809	1.125	.750	.1875				AIEE
48	.01568	.063	.24894	6.485	1.0625	.563	.25				RL-34, RZ-7
49	.01590	.0234	.67929	9.49	1.375	1.000	.125	1.445	.930	.195	MI-50014
50	.01701	.07031	.24192	6.49	1.00	.625	.375	1.070	.555	.445	MI-50094
51	.01703	.0469	.36316	6.97	1.000	.750	.375	1.070	.680	.455	MI-50168
52	.01712	.046875	.36316	7.49	1.125	.750	.250	1.195	.680	.330	MI-50084, RZ-8
53	.01901	.1250	.1521	4.99	.750	.500	1.00	.810	.440	1.110	MI-50110
54	.02011	.0625	.32169	7.57	1.200	.700	.250	1.260	.640	.320	MI-50178
55	.02126	.0313	.67929	8.99	1.250	1.000	.250	1.320	.930	.330	MI-50004, RM-13
56	.02147	.09375	.2290	6.79	1.100	.600	.375	1.160	.540	.440	MI-50177
57	.02310	.0938	.24630	7.18	1.150	.650	.375	1.240	.560	.450	AE-6180, MM-6, MI-50133
58	.02340	.094	.24894	6.485	1.0625	.563	.375				RL-7
59	.02553	.0703	.36316	7.49	1.125	.750	.375	1.195	.680	.455	MI-50034, RZ-4
60	.02685	.0313	.85767	9.99	1.375	1.125	.250	1.445	1.045	.330	MI-50115
61	.02783	.063	.44178	7.980	1.25	.75	.25				RH-17
62-H	.03118	.0469	.66476	9.48	1.375	1.0	.250	1.455	.920	.320	AE-4635,MM-10,MI-50029,RZ-13
63	.03360	.03125	1.07513	10.99	1.500	1.250	.250	1.570	1.170	.330	MI-50039
64	.03407	.0938	.36316	7.49	1.125	.750	.500	1.195	.680	.590	MI-50099
65	.03407	.0938	.36316	7.98	1.250	.750	.375	1.320	.680	.455	MI-50188, RH-44
66	.03578	.0703	.50895	8.48	1.250	.875	.375	1.320	.805	.440	MI-50181
67	.03970	.1641	.24192	8.80	1.500	.625	.375	1.570	.555	.455	MI-50023
68	.04021	.1250	.32169	7.57	1.200	.700	.500	1.260	.640	.570	MI-50179
69-L	.04246	.0625	.67929	9.99	1.500	1.000	.250	1.570	.930	.330	MI-50032,RH-71,RM-4,RL-35
70	.04256	.1172	.36316	8.48	1.375	.750	.375	1.445	.680	.455	MI-50189
71	.05498	.070	.78539	9.476	1.375	1	.375				RL-12
72	.05522	.125	.44178	7.980	1.25	.75	.50				RH-7

*See footnotes at end of table.

(*Table II continued on next page.*)

Table II (continued)

					CORE SIZE			CASE SIZE			
Item No.	Window Area x Iron Area (inches)^4	Iron Area (inches)^2	Window Area (inches)^2	Mean Length of Magnetic Circuit cm.	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	Catalog Number*
73	.06087	.0703	.86590	10.49	1.50	1.125	.375	1.570	1.050	.445	MM-12
74	.06229	.141	.44178	8.992	1.50	.75	.375				WC-1002
75	.06235	.0938	.66476	9.97	1.50	1.0	.375	1.580	.920	.445	AE-5233,MM-11,MI-50026,RL-18,RZ-5,RH-45
76	.06719	.0625	1.0751	11.95	1.750	1.250	.250	1.830	1.170	.320	AE-5387,MM-13,MI-50030,RL-1,RZ-14
77	.07383	.094	.78539	9.476	1.375	1	.50				RZ-16
78	.07681	.0978	.78539	9.23	1.313	1.000	.625				RA-21
79	.08491	.1250	.67929	9.99	1.500	1.000	.500	1.570	.930	.590	MI-50038,RL-40,RZ-19,RH-9,RA-70
80	.08548	.0468	1.8265	14.48	2.0	1.625	.250	2.110	1.525	.330	MM-16,AE-4179, MI-50018
81	.08553	.25	.342119	7.98	1.250	.750	1.0	1.340	.660	1.080	AE-5762, MI-50124
82	.09208	.0938	.98168	10.99	1.563	1.188	.500	1.633	1.118	.590	MI-50058
83	.09322	.211	.44178	10.473	1.875	.75	.375				RL-45
84	.09344	.094	.99401	10.972	1.625	1.125	.375				RH-4
85	.09355	.063	1.48489	12.967	1.875	1.375	.25				RM-6
86	.09675	.219	.44178	9.476	1.625	.75	.50				RM-3
87	.10085	.0938	1.07513	11.49	1.625	1.250	.500	1.705	1.170	.590	MI-50019
88	.10085	.0938	1.07513	11.99	1.750	1.250	.375	1.820	1.170	.455	MI-50098
89	.1072	.1250	.85767	10.99	1.625	1.125	.500	1.705	1.045	.590	MI-50091, AE-5778
90	.11074	.141	.78539	10.972	1.75	1	.375				RH-41
91	.11133	.063	1.76714	13.964	2.	1.50	.25				RL-39
92	.12252	.156	.78539	9.975	1.5	1	.625				RH-68
93	.12737	.1875	.67929	10.96	1.750	1.000	.500	1.820	.930	.590	MI-50190
94-H	.13439	.125	1.07513	11.99	1.750	1.250	.500	1.820	1.170	.590	MI-50035, RH-8, RZ-6, RL-46
95	.13733	.031	4.43013	19.949	2.625	2.375	.25				RH-22
96	.14439	.0938	1.53938	13.99	2.000	1.500	.375	2.100	1.400	.445	MI-50139
97	.14439	.0938	1.53938	13.49	1.875	1.500	.500	1.975	1.400	.590	MI-50087
98	.15507	.156	.99401	11.471	1.75	1.125	.50				RL-41
99	.15507	.039	3.97607	19.203	2.563	2.25	.25				RH-73
100	.16611	.094	1.76714	14.962	2.25	1.50	.25				RL-3
101	.1681	.1563	1.07513	12.49	1.875	1.250	.5000	1.975	1.170	.590	MI-50008
102	.17133	.0938	1.82654	14.49	2.000	1.625	.500	2.100	1.525	.590	MI-50078, RM-11, RL-11
103-L	.17303	.141	1.22718	12.967	2	1.25	.375				RL-6, RH-19, RH-18
104	.18378	.234	.78539	10.972	1.75	1	.625				RH-67
105	.18815	.1750	1.07513	11.39	1.600	1.250	1.000	1.670	1.170	1.110	MI-50090
106	.1924	.125	1.5393	13.97	2.0	1.50	.500	2.110	1.40	.610	MM-14, MI-50055, RM-14
107	.19635	.250	.78539	11.97	2.000	1.000	.500				RA-12, RH-23, RH-79, WC-1003
108	.20159	.1875	1.07513	12.97	2.000	1.250	.500	2.100	1.170	.590	MI-50169, RH-42
109	.20937	.141	1.48489	13.964	2.125	1.375	.375				RA-53
110	.22070	.281	.78539	10.972	1.75	1	.75				RZ-11
111	.22610	.094	2.40528	16.957	2.5	1.75	.25				RH-80
112	.22641	.082	2.76116	16.709	2.313	1.875	.375				RA-6
113	.23164	.156	1.48489	13.466	2	1.375	.50				RL-47
114	.24917	.141	1.76714	14.962	2.25	1.5	.375				RM-7
115	.25761	.328	.78539	10.972	1.75	1	.875				RZ-10
116	.25761	.328	.78539	11.471	1.875	1	.75				RH-59
117	.25924	.125	2.07394	14.962	2.125	1.625	.500				RA-67
118	.2597	.250	1.0386	11.95	1.750	1.250	1.0	1.850	1.150	1.090	AE-7189, MI-50012
119	.28716	.234	1.22718	12.97	2.000	1.250	.625				RA-8, RH-10
120	.29111	.215	1.3540	13.218	2.00	1.313	.625				RA-22
121	.29243	.141	2.07394	15.959	2.375	1.625	.375				RH-56
122	.31113	.313	.99401	10.972	1.625	1.125	1.25				RH-76
123	.32306	.117	2.76116	16.833	2.344	1.875	.50				RH-70, RH-69
124	.33915	.141	.78537	16.957	2.5	1.75	.375				RM-3
125-H	.3396	.125	2.7171	17.95	2.50	2.0	.50	2.652	1.860	.610	AE-4178,MM-17,MI-50017,RL-25
126	.34248	.1875	1.82654	14.49	2.000	1.625	1.000	2.100	1.525	1.110	MI-50020
127	.34515	.125	2.76116	16.957	2.375	1.875	.50				RH-63
128	.36757	.117	3.14159	19.702	2.938	2.000	.250				RA-61
129-L	.3794	.250	1.5174	15.96	2.50	1.50	.50	2.610	1.390	.610	AE-5320,MM-15,MI-50001,RH-39,RZ-20
130	.38485	.2500	1.53938	13.99	2.000	1.500	1.000	2.100	1.400	1.110	MI-50040
131	.39864	.1406	2.83528	18.99	2.750	2.000	.375	2.850	1.900	.475	MI-50056, RM-9
132	.40317	.3750	1.07513	12.99	2.000	1.250	1.000	2.100	1.170	1.110	MI-50036
133	.41726	.281	1.48489	13.218	1.938	1.375	1.				RZ-17
134	.44218	.563	.78539	13.96	2.500	1.000	.750				RA-13
135	.45282	.164	2.76116	16.458	2.250	1.875	.875				RA-71, RM-10, RL-10
136	.45282	.164	2.76116	18.453	2.750	1.875	.375				RA-19
137	.46142	.094	4.90873	21.944	3	2.5	.375				RH-64
138	.48530	.234	2.07394	20.448	3.5	1.625	.25				RH-32
139	.53456	.2500	2.13824	15.99	2.250	1.750	1.000	2.350	1.650	1.110	MI-50131
140	.5468	.125	4.3743	21.90	3.0	2.50	.500	3.152	2.360	.610	MM-18,AE-4180,MI-50031,RZ-22,RL-26
141	.55292	.176	3.14159	19.702	2.938	2.000	.375				RA-60
142	.55312	.313	1.76714	15.959	2.50	1.50	.625				RH-43
143	.557	.375	1.48489	14.961	2.375	1.375	.750				WC-1004
144	.57727	.3750	1.53938	14.99	2.250	1.500	1.000	2.350	1.400	1.110	MI-50088
145	.57727	.3750	1.53938	13.98	2.000	1.500	1.500	2.100	1.400	1.610	MI-50172
146	.64151	.287	2.23521	19.583	3.218	1.687	.375				WC-1005
147	.65039	.438	1.48489	14.463	2.25	1.375	1.				RH-78

*See footnotes at end of table.

(*Table II continued on next page.*)

Table II (continued)

Item No.	Window Area x Iron Area (inches)^4	Iron Area (inches)^2	Window Area (inches)^2	Mean Length of Magnetic Circuit cm.	CORE SIZE			CASE SIZE			Catalog Number*
					Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	
148	.70882	.2500	2.83528	17.99	2.500	2.00	1.000	2.600	1.900	1.110	MI-50041,RA-51,RZ-2,RH-58
149-L	.70882	.2500	2.83528	19.99	3.00	2.00	.500	3.100	1.900	.590	MI-50103, RZ-21, RL-37
150	.71342	.344	2.07394	18.453	3.000	1.625	.500				RA-38
151	.73003	.352	2.07394	20.448	3.5	1.625	.375				RH-33
152	.73513	.234	3.14159	19.702	2.938	2.000	.500				RA-59
153	.74049	.4375	1.69255	16.99	2.625	1.625	..875	2.781	1.468	.995	MI-50113
154	.76576	.156	4.90873	21.944	3.000	2.500	.625				RA-64
155	.76969	.5000	1.53938	15.99	2.500	1.500	1.000	2.600	1.400	1.110	MI-50013, RZ-12
156	.77686	.1250	6.21483	25.99	3.500	3.00	.500	3.688	2.813	.600	MI-50141
157	.78540	1.000	.78539	15.96	3.000	1.000	1.000				RA-14
158	.78540	.250	3.14159	23.939	4	2	.25				RH-30
159	.79482	.253	3.14159	19.203	2.813	2.000	.625				RA-23
160	.80715	.203	3.97607	18.952	2.50	2.25	1.625				RL-38
161	.80806	.2850	2.83528	18.29	2.570	2.000	1.000	2.670	1.900	1.110	MI-50097
162	.84365	.286	2.94983	23.692	4	1.938	.375				RH-55
163	.92123	.383	2.40528	17.456	2.625	1.75	.875				RH-37
164	.97268	.469	2.07394	20.448	3.5	1.625	.50				RH-34
165	.98332	.313	3.14159	19.949	3.000	2.000	.625				RA-42
166	1.02723	.8438	1.21738	15.49	2.500	1.375	1.500	2.630	1.245	1.620	MI-50114
167	1.03541	.375	2.76116	16.957	2.375	1.875	1.500				RA-55
168-H	1.05047	.2500	4.20185	23.99	3.500	2.500	.500	3.688	2.313	.688	MI-50128, MM-19, AE-6100
169	1.05531	.860	1.22718	15.461	2.625	1.25	2.25				RA-52
170	1.06913	.5000	2.13824	17.99	2.750	1.750	1.000	2.850	1.650	1.110	MI-50021
171	1.10446	.625	1.76714	16.957	2.750	1.500	1.000				RA-79
172	1.16500	1.172	.99401	10.848	1.594	1.125	1.25				RH-77
173	1.17810	.375	3.14159	23.939	4	2	.375				RH-29
174	1.17810	.375	3.14159	19.949	3.000	2.000	.750				RA-30
175	1.18521	.219	5.40982	22.942	3.125	2.625	.875				RA-39
176	1.27549	.406	3.14159	19.203	2.813	2.000	1.000				RA-25
177	1.32890	.188	7.06858	26.931	3.75	3	.50				RL-44
178	1.35089	.430	3.14159	21.445	3.375	2.000	.625				RA-9
179	1.35298	.250	5.40982	24.936	3.625	2.625	.50				RA-50
180	1.39284	.938	1.48489	16.957	2.875	1.375	1.25				RA-35
181	1.39574	.215	6.49180	25.687	3.563	2.875	.625				RH-6
182-L	1.41765	.500	2.83528	19.99	3.000	2.000	1.000	3.100	1.900	1.110	MI-50022, RH-20, RH-65, RM-12
183	1.45214	.5625	2.58157	21.99	3.500	2.000	.750	3.688	1.813	.938	MI-50085
184	1.48490	.250	5.93957	22.942	3	2.75	2.				RL-13
185-L	1.56551	.4688	3.33939	22.99	3.500	2.250	.750	3.688	2.062	.938	MI-50064
186	1.57080	.500	3.14159	23.939	4.	2.	.50				RH-31
187	1.65243	.687	2.40528	19.482	3.125	1.750	1.0				WC-1006
188	1.84077	.375	4.90873	23.939	3.500	2.500	.750				RA-41
189	1.86080	.6563	2.83528	18.99	2.750	2.000	1.750	2.850	1.900	1.850	MI-50071
190	1.89746	.7350	2.58157	21.82	3.470	2.000	1.000	3.688	1.813	1.188	MI-50096
191	1.91625	1.00	1.91624	21.99	3.750	1.750	1.000	3.938	1.562	1.188	MI-50028, RH-21
192	1.94525	.313	6.21483	23.692	3.125	2.813	2.				RL-22
193	1.96350	.625	3.14159	20.955	3.25	2.0	1.0				WC-1007
194	1.98800	.500	3.97607	19.949	2.750	2.250	2.000				RA-49
195-H	2.1009	.500	4.2018	23.95	3.50	2.50	1.0	3.688	2.313	1.188	MM20,MI50042,AE5468,WC1008,RH35,RH57,RL14
196	2.11630	1.563	1.3540	15.465	2.563	1.313	2.500				RA-54
197	2.13825	1.00	2.13824	17.99	2.750	1.750	2.000	2.850	1.650	2.110	MI-50132, RH-13
198	2.15003	.438	4.90873	23.939	3.5	2.5	.875				RA-44, RA-58
199	2.16142	.688	3.14159	26.931	4.75	2.	.50				RL-8, RM-1
200	2.21505	.500	4.43013	22.942	3.375	2.375	1.000				RA-63, RH-75, RL-48, RZ-18
201	2.30571	.371	6.21483	27.183	4.	2.813	.625				RA-24
202	2.33314	1.125	2.07394	17.456	2.750	1.625	2.000				RA-73
203	2.34784	.6250	6.23251	24.00	3.625	2.375	1.000	3.813	2.187	1.188	MI-50005
204	2.35620	.750	3.14159	23.939	4.	2.	.75				RH-74
205	2.37041	.438	5.40982	22.942	3.125	2.625	1.75				RA-76
206	2.40529	.250	9.62112	29.924	4.	3.5	1				RA-11
207	2.74890	.875	3.14159	23.939	4.000	2.000	.875				RA-62
208	2.87805	1.500	1.91870	21.96	3.750	1.750	1.500	3.938	1.563	1.688	MI-50148
209	2.90428	1.1250	2.58157	21.99	3.500	2.000	1.500	3.688	1.813	1.688	MI-50062, RA-56, RA-77, RH-60
210	3.06797	.625	4.90873	24.936	3.75	2.5	1.				RH-12
211	3.10742	.500	6.21483	27.99	4.000	3.000	1.000	4.188	2.813	1.188	MI-50044, RA-66, RH-66
212	3.14160	1.0	3.14159	23.939	4.	2.	1.				RL-24
213	3.15140	..7500	4.20185	25.99	4.000	2.500	1.000	4.188	2.313	1.188	MI-50043
214	3.2816	.781	4.2018	24.90	3.750	2.50	1.250	3.938	2.313	1.438	MM-21, MI-50080, RA-2
215	3.68156	.750	4.90873	31.919	5.5	2.5	.50				RH-40
216	3.69162	.594	6.21483	27.183	4.	2.813	1.				RA-26
217	3.76187	.391	9.62112	32.916	4.75	3.5	.625				RL-42
218	3.78596	1.125	3.3653	25.93	4.250	2.250	1.125	4.430	2.070	1.266	AE-6464, RZ-23
219-L	3.9370	.9375	4.2018	24.90	3.750	2.500	1.500	3.938	2.313	1.688	MM-22,MI50100,AE5690,RL15,RZ15,RH-72
220	3.9499	.7656	5.15925	28.99	4.500	2.750	.875	4.688	2.562	1.063	MI-50016
221	3.97609	2.250	1.76716	23.939	4.50	1.50	1.50				RH-38
222	4.38295	1.3125	3.33939	24.99	4.000	2.250	1.500	4.188	2.062	1.688	MI-50093, RA-57, RH-62
223	4.66113	.750	6.21483	27.99	4.000	3.000	1.500	4.188	2.813	1.688	MI-50065
224	4.70925	1.063	4.43013	27.430	4.5	2.375	1				RL-27
225	4.74633	1.719	2.76116	20.448	3.25	1.875	2.500				RA-29

*See footnotes at end of table.

(Table II continued on next page.)

Table II (continued)

Item No.	Window Area x Iron Area (inches)^4	Iron Area (inches)^2	Window Area (inches)^2	Mean Length of Magnetic Circuit cm.	CORE SIZE			CASE SIZE			Catalog Number*
					Outside Diameter	Inside Diameter	Height (inches)	Outside Diameter (inches)	Inside Diameter (inches)	Height (inches)	
226	4.81058	.500	9.62112	31.919	4.5	3.5	1.				RA-68
227	5.07441	.938	5.40982	26.931	4.125	2.625	1.25				RL-49
228	5.16316	2.000	2.58157	23.99	4.000	2.000	2.000	4.188	1.813	2.188	MI-50079
229	5.30145	.750	7.06858	29.924	4.5	3.	1.				RA-48
230	5.52234	1.125	4.90873	25.934	4.0	2.5	1.5				RA-36
231	5.60857	1.266	4.43013	23.440	3.5	2.375	2.25				RH-14
232	5.64506	1.50	4.90873	25.934	4.0	2.5	2.				RA-31
233	5.84597	1.1331	5.15926	29.19	4.563	2.750	1.250	4.750	2.563	1.438	MI-50070
234	6.21206	2.4063	2.58157	26.99	4.750	2.000	1.750	4.938	1.813	1.938	MI-50160
	6.2458	1.00	6.2458	31.93	5.00	3.00	1.00	5.180	2.820	1.140	AE-6379, RL-16
235	6.28320	.50	12.56637	35.908	5.	4.	1.				RA-65
236	6.34394	1.688	3.75825	24.188	3.875	2.1875	2.				RH-3
237	6.44909	1.250	5.15926	29.92	4.750	2.750	1.250	4.938	2.563	1.438	MI-50184
238	6.46097	2.340	2.76116	22.443	3.750	1.875	2.500				RA-18
239	6.71149	1.688	3.97607	26.931	4.500	2.250	1.500				RA-69
240-H	6.8998	.937	7.3637	30.95	4.50	3.250	1.50	4.688	3.062	1.688	MM-23, MI-50081, AE-5737
241	6.90170	1.406	4.90873	29.924	5.	2.5	1.125				RA-74
242	6.99170	1.125	6.21485	29.92	4.500	3.000	1.500	4.688	2.813	1.688	MI-50159, AE-5581
243	7.06860	1.00	7.06858	27.929	4.	3.	2.				RA-1
244	7.51412	.781	9.62112	30.422	4.125	3.5	2.5				RH-50
245	7.65170	1.039	7.36377	29.45	4.125	3.250	2.375	4.313	3.062	2.563	MI-50173
246	8.28968	1.1250	7.36858	31.99	4.750	3.250	1.500	4.938	3.063	1.688	MI-50101
247	8.49646	1.202	7.06858	29.425	4.375	3.	1.75				RA-37
248	9.2046	1.25	7.3637	30.95	4.50	3.250	2.0	4.688	3.062	2.188	MM-24
249	9.28107	1.313	7.06858	30.921	4.75	3.	1.50				RL-20
250	9.62115	1.0	9.62112	35.908	5.5	3.5	1.				RA-20
251	10.21648	.813	12.56637	38.402	5.625	4.	1.				RA-27
252	10.5046	2.50	4.20185	29.99	5.000	2.500	2.000	5.188	2.313	2.188	MI-50069, RA-75
253	10.60290	1.50	7.06858	31.919	5.	3.	1.5				RL-17
254	10.76915	1.250	8.61530	35.91	5.500	3.500	1.250	5.688	3.312	1.438	MI-50185
255	10.82380	2.0	5.40982	28.926	4.625	2.625	2.				RA-17
256-L	10.88	1.312	8.290	32.9	5.000	3.250	1.500				AIEE
257	12.56640	1.0	12.56637	39.898	6.	4.	1.				RL-21, RL-43
258	12.6606	1.8906	6.6966	35.71	5.850	3.10	1.375	6.030	2.920	1.605	AE-5582
259	14.13720	1.125	12.56637	37.903	5.5	4.	1.5				RA-40
260-H	14.2726	1.25	11.4181	36.90	5.250	4.0	2.0	5.438	3.813	2.188	MM-25, MI-50112
261	16.33387	2.750	5.93957	32.916	5.5	2.75	2.				RA-72
262	17.67150	2.500	7.06858	33.914	5.5	3.	2.				RH-11
263	21.9047	1.50	14.6031	43.90	6.50	4.50	1.50	6.687	4.312	1.688	MM-26, MI-50182
264-H	25.13	2.00	12.567	39.9	6.000	4.000	2.000				AIEE
265	43.73652	3.375	12.95894	39.90	5.750	4.250	4.500	5.938	4.062	4.688	MI-50145

* CATALOG NUMBER CODE

VENDOR	CODE LETTERS
Arnold Engineering	AE
Magnetics Inc.	MI
Magnetic Metals	MM
Thomas and Skinner	WC
Westinghouse	RA,RH,RZ,RL, & RM

H- AIEE Proposed Standards - High Gain
L- AIEE Proposed Standards - Low Gain

Note: Different manufacturers may use core boxes of slightly different dimensions for a given size core. To conserve space each size core is listed only once, according to the smallest core case.

Vol.
Pt. 1
1958

area times the core iron area for any design flux density. (For the rating method upon which the charts are based see Appendix II.) The actual product of the iron area and window area is computed by utilizing materials chosen on the basis of the desired dynamic performance, the cost, etc., and by utilizing a core material thickness which is chosen as suitable for the power frequency of operation.

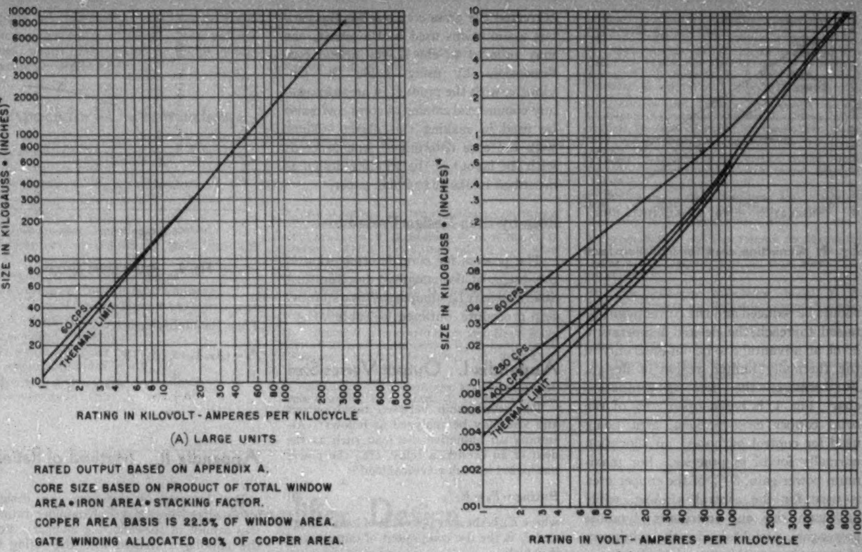
Table II lists some of the available cores from which an actual core is chosen. In Part II techniques are given for computing the number of turns and wire size to meet either a voltage or a load resistance specification. This evaluation is based on equation 1,[11] the transformer voltage equation, and on the available window dimensions.

$$E_s = 4fN_GB \times A_{Fe}(2.54)^2 \times 10^{-8} \qquad (1)$$

Once the core size is known and the wire size chosen, the layer-turns chart, also in Part II, may be used to find the number of turns in the first layer. The winding design continues by using a known number of turns in the first layer, together with a desired total number of turns to enter a chart which gives an equivalent number of layers. From this point, the winding hole size may be determined, while information is given on typical winding capabilities of various toroidal winding machines. With this sequence as a background, the actual use of the design charts will be explained in more detail, together with a step-by-step procedure for undertaking the design.

Core Design

The core determination chart, Fig. 1, is used to determine the core size. Two aspects of this chart deserve special mention. As may be seen from an examination of the chart, one enters this chart on the abscissa axis at the desired maximum rating of the magnetic amplifier divided by the minimum supply frequency. This technique permits the use of a single chart for all frequencies of power supply. For an amplistat circuit with a resistive load, multiply the actual rating by the factor shown in Table III to obtain the rating of a pair of the same reactors in single-phase full-wave single-ended circuits. It may be seen that core sizes listed on the ordinate axis of this chart express the core size in kilogausses multiplied by inches to the exponent four. By employing this method of incorporating flux density into the size figure, the chart may be used with all core materials. By studying the table of dynamic power gains, Table IV, one may choose materials having performance and cost suitable for



(A) LARGE UNITS

RATED OUTPUT BASED ON APPENDIX A.
CORE SIZE BASED ON PRODUCT OF TOTAL WINDOW AREA • IRON AREA • STACKING FACTOR.
COPPER AREA BASIS IS 22.5% OF WINDOW AREA.
GATE WINDING ALLOCATED 80% OF COPPER AREA.
RATING SINGLE PHASE BASIS; FULL WAVE, SINGLE ENDED CIRCUIT.

(B) SMALL UNITS

Fig. 1. Toroidal amplistat core determination chart

the specific application. Obviously, dividing the size, kilogausses times inches to the fourth power, found in the chart by the design flux density, gives the net core size in iron area times window area. The approximate design densities in Table V may be used in establishing the design flux density for the material chosen. From this net core size, one may find the gross core size by dividing by the stacking factor for the thickness of material used. The designer should consult the manufacturer's data for the actual core used, although Table VI gives typical values.

Corrections should be made to the determined gross core size if any of the design aspects deviate from the assumptions used in the preparation of the chart. These items will now be discussed in more detail.

The winding space allocation assumed in the preparation of the chart was based on the study of present design practice. This allocation is shown in Table I. It is interesting to note that, unlike the windings of a cylindrical coil,[7] the space allocation in a toroidal winding is relatively independent of size. This situation arises because of the differences in insulation requirements of small sizes, as well as the economics of manufacture in winding the larger sizes. Of course, the actual space allocations to be used depend upon the winding machine capabilities and the economics of the particular manufacturing operation. The gross core size required will vary inversely with the per cent of the total window space allocated to copper. This window space allocation may be influenced by the insulation requirements, as well as by the winding machine capabilities. As the required hi-pot level is increased, more of the window space must be allocated to the necessary insulation. Fig. 2 gives approximate factors by which the gross core size should be divided to correct for insulation level. The core determination chart is based on designs having temperature rises of 55 C. For high-temperature materials the size

Table III. Rating Factors

Circuit	Factor	Number of Cores Required
Single-winding	3.4	1
Three-phase bridge	0.30	6
Push-pull	approximately 5	depends on connection

Table IV. Dynamic Power Gain or Figure of Merit

P_{rated}, Volt-Amperes	P_{rated}, Watts Per Kc	D_P w/w Per Cycle	f, CPS	Core Material	Rectifier Material
121.5	84.45	2,200	1,440	Deltamax*	germanium
27	18.75	13,000	1,440	Supermalloy*	
3.10	2.15	2,400	1.440	Supermalloy	
3,380	8,450	190	400	silicon iron	selenium
2,725	6,820	220	400	silicon iron	
418	1,160	600	360	Deltamax	germanium
104.5	249	90	420	silicon iron	selenium
5	11.9	3,300	420	Supermalloy	germanium
0.067	0.158	100	400	Supermalloy	selenium
56,100	936,000	940	60	silicon iron	selenium
5,620	93,700	460	60	silicon iron	
275	4,580	530	60	silicon iron	selenium
2.18	36.4	530	60	Deltamax	germanium

* Arnold Engineering Company.

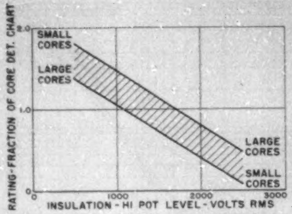


Fig. 2. Correction chart for insulation level

should be reduced and the current density raised to reach the desired temperature level, an advantage to be obtained only in the thermally limited region of design. Furthermore, the core determination chart, Fig. 1, is based upon 20% of the total copper cross-sectional area being used for control purposes; an allocation normally found in practice. For maximum power gain, 50% of the copper area is used for the control winding, while minimum size and maximum dynamic power gains are obtained with a minimum of space given to the control winding.

In thermally limited designs, the minimum space just mentioned is established by the thermal current capacity of the control conductor and it results in approximately 100 divided by K_{AT}, per cent of the total copper area given to the control winding. The gross core size should be corrected for the copper allocation by being multiplied by 80 divided by the actual per cent of copper to be allocated to the gate. Having corrected the gross core size for the various assumptions used in the chart, one may proceed to choose the actual core dimensions by using Table II. Obviously, with the product of areas known, any commercial catalog of cores could also be used in making this choice. Thus, with the core determined, one proceeds with the layout of the winding design as described in Part II of this paper.

Step-by-Step Design Procedure

The process for core determination in the design of the magnetic amplifier just described may be summarized as a step-by-step procedure, outlined in Table VII.

Appendix I. Output Versus Size

The relationship between rated output and size may be analyzed as follows: Assuming an inductive d-c load such as the field of an exciter, a relay, etc., the power dissipated in such a typical load is

$$\text{Rating} = P_o = E_L I_o \qquad (2)$$

where E_L is the average value of load voltage and I_o is the d-c component of output, and in which

$$E_L = E_o \times \frac{R_L}{R_o} = E_o \times \frac{n}{n+1} \qquad (3)$$

or

$$E_L = A_{Fe} s_f B_s (4fN_G) \times 10^{-8} \frac{n}{n+1} \qquad (4)$$

and

$$I_o = I_G k_f \qquad (5)$$

$$= \frac{A_w \times J \times G\, k_f w_f}{N_G} \qquad (6)$$

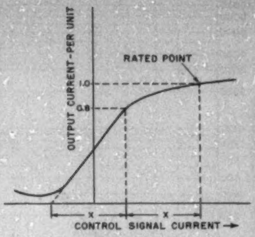


Fig. 3. Rating point diagram

Thus, combining equations,

$$P_o = (A_{Fe} A_w) \times (B_s J) \times \frac{n}{n+1} (s_f w_f G\, k_f) \times 4f \times 10^{-8} \qquad (7)$$

Appendix II. Method of Rating

For purposes of preparing this design information the magnetic amplifier rating was defined in the following manner: For the impedance limited case the rating is taken with impedance match, while in the thermally limited case the rating is taken at a temperature rise of 55 C. Resistive loading is used in both cases with the rated point determined to meet the geometry of Fig. 3, which shows the rated point so placed that the output is down 20% when the signal is moved halfway to cut-off. This places the point of rating beyond the knee of the characteristic, representing a usable choice for maximum output. For the resistance-limited case the rating point is determined graphically from Fig. 3. In the thermally limited case the load is adjusted to place the thermal limit at the one-per-unit output

Table V. Approximate Design Densities

Core Material	Design Flux Density, Kilogausses
Deltamax or Orthonol, etc.	14.5
Mo-permalloy	6
Mumetal	5.5
Nicaloi	12.5
Supermalloy	6.75
Audio-grade silicon iron	15.5 to 17.5

Table VI. Stacking Factor, Tape-Wound Toroids

Tape (Material) Thickness, Thousandths of an Inch	Stacking Factor
1	0.70
2	0.80
4	0.85
10	0.93

Table VII. Design Procedure

Step I, Core Size Determination

Steps	Method
A. Convert frequency to kc	$f = \frac{\text{cps}}{1{,}000}$ kc
B. Divide desired rating by frequency	$\frac{\text{maximum output (equation 2)}}{\text{minimum } f \text{ in kc}}$
C. Convert single-winding 3-phase and push-pull circuits to equivalent	use rating factors, Table III
D. Find size from core determination chart	use Fig. 1
E. Choose a core material and thickness	use Table IV
F. Establish a design flux density	use Table V
G. Find net core size, inches^4	$\frac{\text{core size from step D}}{\text{flux density, step F}}$
H. Obtain stacking factor	use Table VI
I. Find gross core size, inches^4	$\frac{\text{net core size, step G}}{\text{stacking factor, step H}}$
J. Apply appropriate corrections to core size if design varies from the usual	
1. Winding machine capability	see Table I and Part II
2. Insulation level	see Fig. 2 and Part II
3. Temperature rise design	multiply by $\sqrt{55\text{ C}/\theta\text{ C}}$
4. Copper allocation	multiply by $\frac{80}{\text{per cent copper in gate}}$
K. Using corrected gross core size from step I or step J, find core dimensions	use Table II or vendor's catalog; consult list of references

of Fig. 3. The method described to define the rating is one (out of many other possible ones) which happened to be useful in preparing the design material.

Appendix III. Nomenclature

A_{Fe} = cross-sectional area of iron (net), square inches
B_s = saturation flux density, gausses
D_p = dynamic power gain[11]
E = voltage, average
f = frequency, cps (cycles per second)
G = fraction of copper in gate winding, usually 0.8
I_G = gate current rms, amperes
I_o = average output current, amperes
J = current density, amperes per square inch
k_f = a form factor
K_{AT} = ampere-turn gain
N_G = gate turns per core
n = matching index, $n=R_L/R_G$
P_o = output power, see equation 2
R = resistance, ohms
s_f = stacking factor
w_f = winding factor
w/w = power gain in watts per watt
θ = temperature rise, C

Subscripts

G = gate
L = load
o = output

References

1. An Analysis of Optimum Core Configuration for Magnetic Amplifiers Using a Simplified Approach, B. D. Bedford, C. H. Willis, G. C. Dodson. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 62–70.

2. Effect of Core Material on Magnetic Amplifier Design, A. O. Black. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 4, 1948, pp. 427–35.

3. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, C. W. Lufcy, H. H. Woodson. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 220–26.

4. Rules of Similitude for Magnetic Amplifier Systems, L. A. Finzi, H. L. Durand. *Proceedings*, National Electronics Conference, vol. 7, Oct. 1951, pp. 498–514.

5. Influence of ID/OD Ratio on Static and Dynamic Magnetic Properties of Toroidal Cores, R. W. Roberts, R. I. Van Nice. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 599–607.

6. Analysis and Design of Self-Saturable Magnetic Amplifiers, Sidney B. Cohen. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 39, Sept. 1951, pp. 1009–20.

7. The Transductor Amplifier (book), Ulrik Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947.

8. Electronic Transformers and Circuits (book), Reuben Lee. John Wiley & Sons, Inc. New York, N. Y., second edition, 1955.

9. Some Notes on the Design of Magnetic Amplifiers, A. S. Fitzgerald. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 244, Nov. 1947, pp. 323–62.

10. The Magnetic Amplifier (book), J. H. Reyner. Stewart and Richards, London, England, Feb. 1951.

11. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., 1955.

12. Designing Saturable-Core Reactors for Specific Use, Carl Helber. *Electronic Industries and Electronic Instrumentation*, New York, N. Y., Dec. 1947, pp. 1–13.

Magnetic-Amplifier Design

Part II—The Windings

R. E. ANDERSON
ASSOCIATE MEMBER AIEE

Synopsis: This development of a practical approach to the design of the toroidal magnetic amplifier is based on synthesizing the design from design charts, thus avoiding the necessity for having available detailed data from previous equipment. Starting with a known power supply frequency, together with a desired gain and time constant, the load voltage and current are utilized in determining the core. With the core specified, additional charts and techniques are presented for the winding layout design. The usual heat transfer and resistance computations serve as a check on the initial design synthesis, and these complete the design process. A straightforward and practical step-by-step design procedure is outlined. This paper, the second of two on the subject, is concerned with the winding layout and the completion of the design. Part I,[1] presented the design approach and the method of core determination. Thus, information is given making possible the practical and effective design of the magnetic amplifier having toroidal construction.

WIDESPREAD INTEREST in the use and application of the magnetic amplifier has led to the need for an effective and practical method for the design of this component. For the purpose of supplying information to meet this need, charts have been prepared and data collected which will permit the design of toroidal magnetic amplifiers without reference to previous designs. The engineering problem is treated in Part I, together with the design procedure, which is based on, first, determining the core product of window area times iron area by using power rating and, second, by laying out a winding design utilizing the chosen core. This paper continues the design technique by making the turns computation and by making the winding layout design. Techniques for verifying the design synthesis are outlined, together with a step-by-step design procedure to enable the reader to achieve a clear-cut solution to his design problem.

Continuation of the Design Technique

Having chosen a suitable core by any available means such as the method described in Part I, the design process may be continued by computing the number of turns required for the gate winding. If the power supply voltage is specified, equation 1 may be utilized in calculating the gate turns.

$$N_G=\frac{E\times10^8}{4.44fA_{Fe}B_s(2.54)^2} \tag{1}$$

On the other hand, if one is working in the resistance-limited region or if the load impedance specifies that the gate winding should meet a certain required resistance, equation 2 may be utilized in determining the number of gate turns.

$$N=\sqrt{\frac{R_GA_{Cu}}{\rho\ MLT}} \tag{2}$$

When equation 2 is used to determine the gate turns, the appropriate power supply voltage may be determined from equation 1. Since approximately 23% of the window area is allocated to copper, and since one normally allocates 80% of the copper space to the gate winding, the cross-sectional area of the gate-winding conductor may be computed. This is done by dividing the available area by the number of gate turns. Obviously, in making this computation the same space allocation should be applied as was used to determine the appropriate core. Having chosen the core, the turns, and the wire size, the designer may proceed to the winding design charts. These additional charts are used to complete the winding

---

Paper 58-39, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted September 18, 1957; made available for printing November 6, 1957.

R. E. Anderson is with the General Electric Company, Schenectady, N. Y.

The author wishes to thank Dr. H. F. Storm for his provident counseling and helpful guidance which were so essential to the preparation of this material. Thanks also go to the author's many colleagues at the General Electric Company who graciously supplied their design data on magnetic amplifiers which exhibited production successes.

layout and to compare design dimensions with winding machine capabilities.

Starting with the layer-turns chart, Fig. 1, the winding layout is begun by evaluating the number of turns in the first layer. This chart takes into account the fact that the lack of precision in a typical toroidal winding machine results in only 95% of the inner perimeter being used for winding. In the preparation of the chart it has been assumed that a standard single coating thickness of insulating enamel is used on wire sizes AWG *41* and smaller, while a standard double-coating thickness of insulating enamel is employed on wire of size AWG *40* and larger. For other insulation thicknesses the proper correction should be applied. With a precision winding, an indication of the turns in the first layer would completely define the winding geometry. Although this is not obvious, it arises from geometrical considerations which yield the fact that in a precision winding, each additional layer has a fixed number, Δ, fewer turns than the previous layer, as will be analyzed in Appendix I. This loss in turns per layer is independent of the relative hole-to-wire diameter and approaches $\Delta = 2\pi$ in a multilayer winding. Although in actual practice most winding machines do not build a precision winding, it is possible to express the actual random-wound coil in terms of an equivalent number of fictitious layers. Thus, by applying appropriate factors to the precision winding, it is possible to obtain an accurate expression for practical winding machine performance. Since the layers may intermesh, rather than building on top of one another as assumed in a precision winding, the winding factor may approach unity. In Figs. 2 and 3 the intercept of the *X*-axis value of the number of turns in the first layer, with the *Y* value for the total number of turns in the winding, locates the design point. This point locates the number of layers in the winding and determines the approximate percentage of available copper which has been utilized by the windings. For windings having less than 360 degrees traverse, the number of turns in both the total winding and the first layer should be multiplied by 360 degrees divided by traverse degrees before entering the chart. With this multiplying technique the correct number of layers and correct space allocation may be read directly. Knowing the number of layers permits the evaluation of coil build, mean length of turn, and winding resistance in the usual manner as given by equations 3, 4, and 5.

$$t_w = \frac{L \times D_c}{0.95} \tag{3}$$

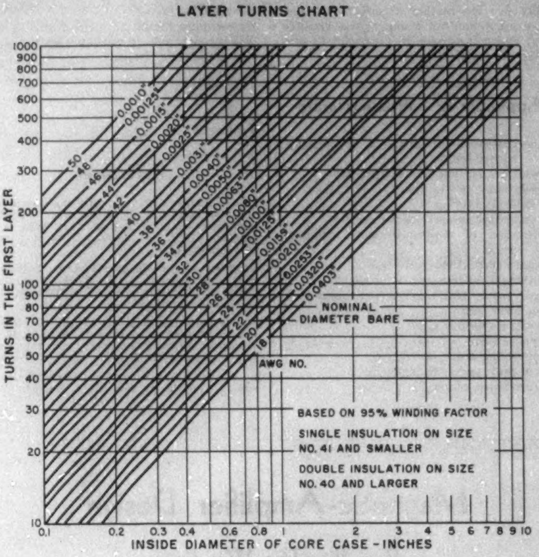


Fig. 1. Layer turns chart

$$MLT = 2[(t_s + t_t + t_w/2)\pi + h + F] \tag{4}$$

$$R = \rho \times \frac{MLT \times N}{12{,}000} \tag{5}$$

As a final check on the design geometry, the hole determination charts, Figs. 4 and 5, may be used. These charts, based on the same space allocations and wire insulation previously summarized, are used to estimate the hole remaining in the toroid after winding.

The intersection of the desired number of layers line with the chosen wire-size line in the rectangular portion of the chart represents the starting point for using these graphs. From this point one follows the rectilinear co-ordinates leftward to the triangular portion of the chart. By following the triangular coordinates in a downward and leftward direction to the intersection of the appropriate line for hole diameter before winding, the design point may be located. The scale on the left-hand side opposite the design point shows the hole remaining after winding while typical space allocation boundaries are shown for reference. The hole dimensions may be compared with the various winding machine capabilities shown in Table I. This table of winding machine data gives dimension limits and appropriate wire sizes for some of the commercially available toroidal winders.

After the gate winding has been laid out, the design proceeds to the control winding. The control turns are chosen to produce a specified current gain or to achieve a specified winding resistance. Probably the best way to determine the necessary number of turns for the control winding is to use data obtained from a core-testing equipment.[2] However, if such information is not available, a theoretical equation or a modification of this equation, which has been given by Storm,[3] may be used as shown in equation 6:

$$N_c I_c = \frac{H_c l_{Fe}}{0.4\pi} \tag{6}$$

Once the control turns have been chosen from core test data, from equation 2 or from equation 6, and with the available copper area known, the winding layout proceeds by the same steps previously discussed for the gate winding.

This process completes the design synthesis and design calculations. Nevertheless, it is advisable to check these computations to be sure that the design fully satisfies the needs and specifications involved. If the design is resistance-limited, i.e., comes from the lower left-hand portion of the core determination chart, Fig. 1 of Part I, the condition of impedance matching may be checked by a comparison between the calculated winding resistance plus rectifier forward drop against the actual load impedance. On the other hand, if the design is in the thermally limited region the use of Fig. 1 of Part I establishes the approximate heat-dissipating capacity; for a more accurate calculation the entire design rating may be checked by using the heat transfer information given in references 4 through 7. Needless to say, the actual winding space should be compared with the capabilities of the winding machines available for the job. The design process just described in detail is summarized in step-by-step form as shown in Table II.

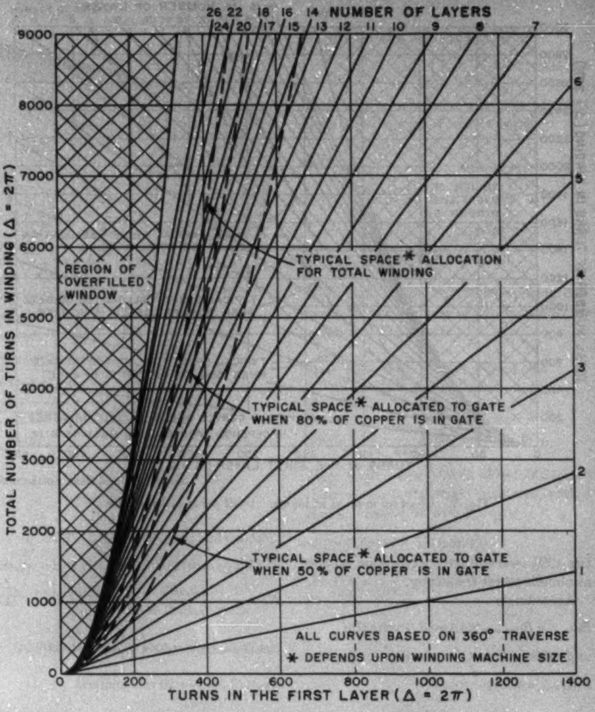


Fig. 2. Chart for number of layers: large cores or small wire

Design Example

As an example, consider the design of a center-tap full-wave single-ended magnetic amplifier to meet the following performance goals:

Power output (rated point; see Fig. 3, Part I): 14.5 va (volt-amperes), at room temperature with nominal voltage and frequency.

Load: 52.5 ohms resistive.

Power frequency: 60 cps (cycles per second) ±5%.

Maximum signal current: 5.5 milliamperes.

Design objective: high dynamic power gain with reasonable weight and size.

With these specifications as the starting point, the design may be calculated by following the steps given in Table VII of Part I, and Table II of the present paper. Accordingly, starting with Table VII, Part I, convert the minimum frequency to kilocycles:

Step I-*A*

$$f=\frac{57}{1,000}=0.057 \text{ kc}$$

With the specified 10% overvoltage, the maximum output becomes 14.5 $(1.1)^2$ = 17.6 va.

Step I-*B*

$$\text{output/kc}=\frac{17.6}{0.057}=308 \text{ va/kc}$$

Step I-*C:* From Fig. 1, Part I, size = 3.90 kilogausses × inches^4

Step I-*D:* To meet the design objective of high dynamic power gain would imply a permalloy type of core material while the size and weight objective could best be met by a silicon steel. A reasonable compromise between these materials (50% nickel, 50% iron core) will be chosen to approach both objectives. The thickness suitability depends upon the power frequency; for this example, at 60 cps, 4-mil or thinner tape may be used. For instance, in the following calculations consider 2-mil tape thickness.

Step I-*E:* From Table V, item 1, Part I, B = 14.5 kilogausses.

Step I-*F:* Determine the net quartic volume as

$$NQV=\frac{3.90}{14.5}=0.269 \text{ in}^4$$

Step I-*G:* From Table VI, Part I, s_f = 0.8.

Step I-*H:* Now the gross quartic volume may be determined:

$$GQV=\frac{NQV}{s_f}=\frac{0.269}{0.8}=0.336 \text{ in}^4$$

Step I-*I:* To obtain the high dynamic power gain objective, allocate most, e.g., 99%, of the copper space to the gate. Then

$$GQV_a=\frac{80}{99}\times 0.3362=0.2717 \text{ in}^4$$

Step I-*J:* From Table II, Part I, choose item 118.

Step II-*A:* First determine the appropriate line voltage

$$V_{\text{load}}=\sqrt{PR_L}=\sqrt{14.5\times 52.5}=27.6 \text{ volts d-c nominal}$$

With this nominal load voltage, the appropriate line voltage may be obtained by taking into consideration the gate-winding resistance (see Appendix II), the after-saturation inductance, and the form factor.

$$V_{\text{line}}=27.6\times 1.22\times\frac{1}{0.875}\times 1.11=42.7 \text{ volts rms nominal}$$

Of course, the turns are designed to handle

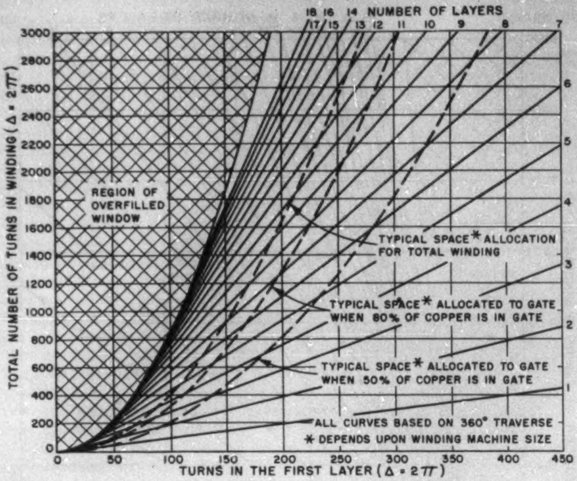


Fig. 3. Chart for number of layers: small cores or large wire

the 10% overvoltage specification at the minimum power frequency.

Accordingly

$E_{\text{line hi}} = E_{\text{line nom}} \times 1.1 = 47.0$ volts rms

and adding 1 volt for rectifier drop gives

$E_{\text{line max}} = 48.0$ volts rms

Then

$$N_G = \frac{E \times 10^8}{4.44 f A_{Fe} B_s (2.54)^2} \quad (1)$$

$$A_{Fe} = h \times F \times s_f = 1 \times \left(\frac{1}{4}\right) \times 0.8 = 0.200 \text{ in}^2$$

$$A_{Fe} \times (2.54)^2 = 1.29 \text{ centimeters}^2$$

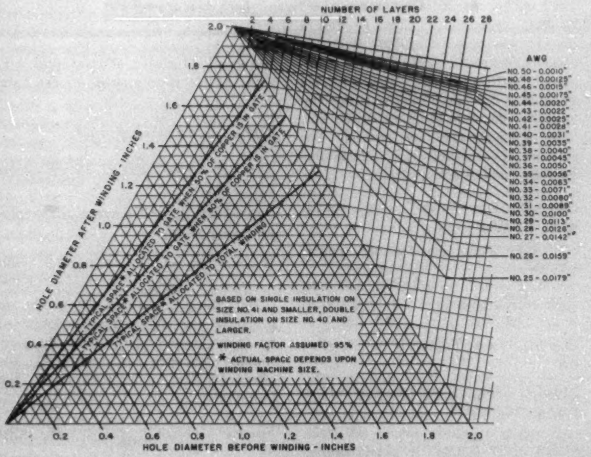


Fig. 4. Toroidal amplistat hole determination chart for small cores

$$\Phi = B_s A_{Fe} = 14{,}500 \times 1.29 = 18{,}710 \text{ maxwells}$$

$$N_G = \frac{48.0 \times 10^8}{4.44 \times 57 \times 18{,}710} = 1{,}014 \text{ turns}$$

Step II-*B*

$$A_{Cu} = A_w \times 0.225 = 1.0386 \times 0.225 = 0.234 \text{ in}^2$$

Step II-*C*

$$A_G = A_{Cu} \times 0.8 \times J_i = 0.234 \times 0.8 \times \frac{0.99}{0.8} = 0.231 \text{ in}^2$$

Step II-*D*

$$Ac_G = \frac{A_G}{N_G} = \frac{0.231}{1{,}014} = 228 \times 10^{-6} \text{ in}^2$$

Therefore try AWG *25* = 0.0179-inch diameter or 0.0206-inch insulated with a double coating of Formex* wire enamel.

$Ac_G = 252 \times 10^{-6}$ in² (actual)

$$I_L = \sqrt{\frac{P}{R}} = \sqrt{\frac{17.6}{52.5}} = 0.579 \text{ ampere d-c in load}$$

$$I_G \text{ per winding} = I_L k_f = 0.579 \times 0.86 = 0.498 \text{ ampere rms in gate}$$

$$J = \frac{I}{Ac_G} = \frac{0.498}{252 \times 10^{-6}} = 1{,}976 \text{ amperes/in}^2$$

Step III-*A*: From Fig. 1, at ID = 1.15, T_1 = 165 turns/layer.

Step III-*B*: Use standard insulation.

Step III-*C*: Use 360 degrees traverse.

Step III-*D*: From Fig. 3, at T_1 = 165 and N = 1,014, $L = 6.9 \cong 7$ layers.

Step III-*E*: Also note from Fig. 3 that available space is adequate.

Step IV-*A*: From Fig. 4 with ID = 1.15 and seven layers of AWG *25*, hole = 0.850 inch with gate winding applied before coil insulation is applied. Since the wraps of coil insulation must insulate the outer perimeter of the winding, the "bunching" on the inside will result in an insulation thickness of nearly 0.040 inch or a hole decrease of 0.080 inch.

Then, the hole after insulating will be 0.850 − 0.080 = 0.770 inch.

Step IV-*B*: Also note from Fig. 4 that available space is adequate.

Step IV-*C*: From Table I note that the following winding machines could be used for this job: Boesch standard, Harder, Micafil *RW-I*, or Rex *DB-2*.

$$MLT = 2\left[\left(t_s + t_w + t_i + \frac{t_G}{2}\right)\pi + A + F\right] \quad (4)$$

$$= 2[0.15\pi + 1.11 + 0.35] = 3.86 \text{ inches} = 0.322 \text{ feet}$$

$$\text{Length} = N\,MLT = 1{,}014 \times 0.322 = 326 \text{ feet per winding}$$

---

*Registered trade-mark, General Electric Company.

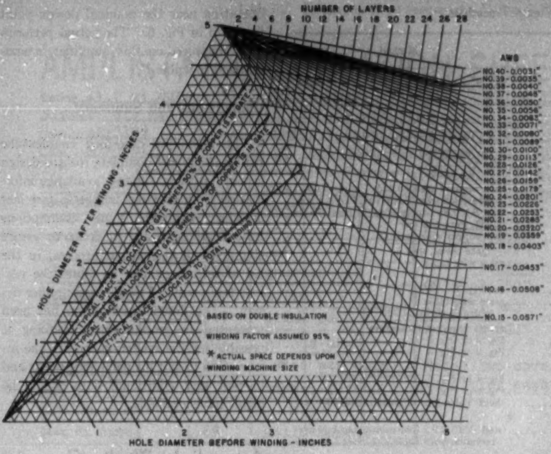


Fig. 5. Toroidal amplistat hole determination chart for large cores

$$R = \rho MLT = 33.0 \times 0.326 = 10.8 \text{ ohms} \quad (5)$$

$$\pm 6\%^{*} = 10.2 \text{ to } 11.4 \text{ ohms}$$

$$W = g\ MLT = 0.986 \times 0.326 = 0.322 \text{ pound}$$

$$R_{hot} = 1.19 \times 11.4 = 13.6 \text{ ohms maximum}$$

$$I^2 R_G = (0.498)^2 \times 13.6 = 3.4 \text{ watts}$$

Step V-*A*

$$A_C = A_{Cu} \times 0.2 \times J_s = 0.234 \times 0.2 \times \left(\frac{1-0.99}{0.2}\right) = 0.00234 \text{ in}^2$$

Step V-*B:* Determine the control turns from the dynamic coercive force

$$N_C = \frac{AT_c}{I_C}$$

$$AT_c = \frac{H_c l_{Fe}}{0.4\pi} = \frac{0.18 \times 11.99}{0.4\pi} = 1.72 \text{ ampere-turns} \quad (6)$$

$$N_C = \frac{1.72}{5.5 \times 10^{-3}} = 312 \text{ turns}$$

Note that at this point the exciting current, I_{ex}, for a-c test of the core may be computed as follows:

$$I_{ex} = \frac{AT_c}{N_G} = \frac{1.72}{1{,}014} = 1.69 \text{ milliamperes}$$

Step V-*C:*

$$Ac_C = \frac{A_C}{N_C} = \frac{0.00324}{312} = 7.49 \times 10^{-6} \text{ in}^2$$

Step V-*D*: Use AWG *40* = 0.0031-inch diameter or 0.0039-inch insulated with a double coating of wire enamel.

$$Ac_C = 7.5 \times 10^{-6} \text{ in}^2 \text{ (actual)}$$

$$J = \frac{I}{Ac_C} = \frac{5.5 \times 10^{-3}}{7.5 \times 10^{-6}} = 733 \text{ amperes/in}^2$$

---

* Each designer should use the tolerance appropriate to his particular manufacturing operation.

Table I. Toroidal Winding Machine Data

	Dimension Limits on Core Plus Winding				Wire Size						
Item	Max. O.D. (Inches)	Min. I.D. (Inches)	Max. h (Inches)	Random or (R) Precision (P) Winding Type	Max. (Inches)	Min. (Inches)	Shuttle Capacity	Shuttle Size	Winding Rate Turns/minute	Machine Manufacturer	Model
1.				P	0.010	0.0015				Universal	U-14A
2.	1 1/2	3/16	1 1/2	R	0.0226	0.0031	0.14 in^3	1/8x7" dia	400	Harder	M
	1 1/2	1/4	1 1/2	R	0.0226	0.0031	0.40 in^3	3/16x7" dia	400		
	1 1/2	5/16	1 1/2	R	0.0226	0.0031	0.79 in^3	1/4x7" dia	400		
	1 1/2	1/2	1 1/2	R	0.0226	0.0031	1.92 in^3	3/8x7" dia	400		
3.	1 1/2-1/2	5/16-3/8	13/16-1 1/16	R	0.008	0.0025	0.0067 lbs	3"S	1000	Boesch	Std., Autom.
	1 1/2-3/4	3/8-1/2	13/16-1 1/2	R	0.0159	0.0025	0.018 lbs	3"	1000		
	2 1/2-1	3/8-5/8	3/4-1 7/8	R	0.0159	0.0025	0.029 lbs	4"S	750		
	2 5/8-1	1/2-11/16	3/4-1 3/4	R	0.0253	0.0031	0.075 lbs	4"	750		
	4 7/8-1	5/8-3/4	1 3/4-1 7/8	R	0.0253	0.0031	0.10 lbs	6"S	500		
	5-2	13/16-7/8	1 5/8-1 7/8	R	0.0320	0.0040	0.30 lbs	6"	500		
4.	1	1/8	1/2	R	0.004	0.0015	Unknown		300	Boesch	Subminiature Auto.
5.	3	5/32	2	R	0.0113	0.002	330 ft-0.011	1/8"	860	Harder	L-2
	3	7/32	2					13/16"	860		
	3	9/32	2					1/4"	860		
	3	13/32	2					3/8"	860		
6.	4	9/16	2	P	0.0226	0.0050	1 3/4 oz	6" dia	200	Micafil	RW-1
7.	4	1/2	1		0.0159	0.002	300 ft of AWG 40	1/8x1/8 in	1200	Arnold	AWC-4
	4	3/16	1/2		0.0159	0.002	700 ft of AWG 40	3/16x3/16 in	1200		
	4				0.0159	0.002	1240 of AWG 40	1/4x1/4 in	1200		
8.	5 29/32	2	4	R	0.0571	0.0159	4.8 in^3	Large	60	Rex	DB-2 Modified
	5 29/32	1	2 3/4	R	0.0403	0.0100	3.1 in^3	Medium	75		
	5 29/32	19/32	2 3/4	R	0.0253	0.0063	1.0 in^3	Small	100		
9.	6"	15/16	1/4	P&R	0.0453	0.008	6 1/4 oz.		300	Micafil	RW-II
	6"	1	1		0.0453	0.008	6 1/4				
	6"	1 3/16	2		0.0653	0.008	6 1/4				
10.	6	1 1/8	2 1/2	P&R	0.040	0.012	0.25 lbs.	Ring magazine	380	Rex Rheostat	Type DB-intermediate
11.	7.9	1 5/8	1/4	P&R	0.108Al 0.086 Cu	0.0126	4 1/2 lbs.	-	150	Micafil	RW-III
	7.9	1 3/4	1								
	7.9	1 7/8	2								
	7.9	2 1/8	4 1/2								
12.	23	2 7/8	10	R	0.0128	0.010		1"x1/2"	100	Universal	U-20

All figures are only approximate and will vary with actual manufacturing techniques applied. Machine manufacturers should be contacted for specific performance data.

Table II. Step-by-Step Design Procedure

Steps	Method	
I. Core determination	See Part I of this paper	
II. Wire and turns computation		
A. Compute gate turns		
1. For fixed line voltage E, rms	$N=\frac{E\times 10^8}{4.44fA_{Fe}B_s(2.54)^2}$	(1)
2. For a specified gate resistance R_G	$N=\sqrt{\frac{R_C A_{Cu}}{\rho\ MLT}}$	(2)
B. Find total copper area	core window area $\times 0.225$: apply step J1 correction from Table VII, Part I, if any	
C. Find gate copper area	step B $\times 0.8$: apply step J4 from Table VII, Part I, if any	
D. Find gate conductor area	$\frac{\text{gate copper area, step C}}{\text{gate turns, step A}}$	
E. Use conductor area to find wire size	use a wire table	
III. Winding design		
A. Find turns in first layer	use Fig. 1	
B. Correct for unusual insulation (if used)	use a wire table	
C. Consider traverse if less than 360 degrees	$\frac{\text{multiply turns by 360}}{\text{actual traverse in degrees}}$	
D. Find the number of layers	Fig. 2 or Fig. 3	
E. Check space utilization	Fig. 2 or Fig. 3	
IV. Design checks		
A. Find winding hole	Fig. 4 or Fig. 5	
B. Check space utilization	Fig. 4 or Fig. 5	
C. Compare with winding machine capabilities	check Table I or references 8 and 9	
V. Control windings		
A. Find control copper area	(step IIB) $\times 0.2$ (one minus step I-J4 corrections, from Table VII, Part I, if any)	
B. Choose turns		
1. For fixed control current	$N_C=\frac{H_c l_{Fe}}{0.4\pi I_C}$	(6)
2. For specified control winding resistance R_C	$N_C=\sqrt{\frac{R_C A_{Cu}}{\rho\ MLT}}$	(2)
C. Find control conductor area	$\frac{\text{control copper area } A}{\text{control turns } B}$	
D. Use conductor area to find wire size	use a wire table	
E. Proceed with winding design	see step III	
F. Repeat design checks for control winding	see step IV	
VI. Check heat transfer or impedance matching	See references 4 through 7	

Step V-*E:* From Fig. 1, with ID = 0.77 inch and AWG *40*, $T_1 = 590$ turns per layer. Thus this winding could be wound in less than one layer. Frequently, however, it is desirable to leave some of the traverse for other windings; for example, bias, feedback, stabilizing, or extra control windings. Methods for handling such a situation will follow by assuming that the control winding traverse is limited to 50 degrees.

Then, the turns in the first layer T_1 are

$$T_1 = 590\times\frac{50}{360} = 82 \text{ turns}$$

For finding layers use Fig. 2 with $82 \times 360/50 = 590$ turns per layer and with $312\times 360/50 = 2246$ turns. Thus, $L = 3.9 \cong 4$ layers and, from Fig. 4, the hole = 0.73 inch.

Note that now the total space utilization may be checked by locating the intercept of this final hole after winding, 0.73 inch, with the initial case ID of 1.15 inch. At this point it is seen that typical space utilization has been achieved.

Continuing with the control winding calculations:

$$MLT = 2\left[\left(t_s+t_w+t_1+\frac{t_C}{2}\right)\pi + A + F\right] \quad (4)$$

$$= 2[0.21\pi + 1.46] = 4.24 \text{ inches} = 0.353 \text{ foot}$$

$$\text{Length} = N\ MLT = 312\times 0.353 = 110.2 \text{ feet}$$

$$R = \mathrm{P}\ MLT = 1101\times 0.1102 = 121.3 \text{ ohms}$$

$$W = g\ MLT = 0.0298\times 0.1102 = 0.00328 \text{ pound}$$

$$I^2R = (5.5\times 10^{-3})^2\times 121.3\times 1.19\times 1.18 = 5.15\times 10^{-3} \text{ watts}$$

These design calculations are summarized in the calculation sheet shown in Table III. Of course, a heat transfer check similar to that used in transformer design should be undertaken at this point as a final step. With such information available from the literature and with the actual result depending upon the particular mechanical construction used, the heat transfer calculations will not be shown. It may be mentioned that in a construction where the two cores are mounted on top of one another so that approximately half of each surface is available for cooling, the hot spot rise will be nearly 55 degrees centigrade at rated output.

A center-tapped full-wave self-saturating magnetic amplifier was constructed in accordance with these calculations. The control characteristic of this amplifier operating near the nominal design point is shown in Fig. 6. The actual performance measured on this amplifier is summarized in Table IV.

Conclusions and Summary

The technique described enables the reader to prepare complete reactor design layouts without having available information from previous magnetic amplifier designs. Although some assumptions have been necessary in order to assure simplicity of the design process, in the list of references may be found the various fine points of magnetic amplifier design. The design information has been presented in graphic form, making it possible for the reader to modify the information to suit his particular needs and to plot any additional data which he might have available.

Appendix I. Winding Geometry

The turns in a layer of mean diameter D is given by

$$T = \frac{\pi D}{D_c} \quad (7)$$

Now consider the difference Δ in turns per layer found in adjacent layers

$$\Delta = T_{n-1} - T_n \quad (8)$$

$$\Delta = \frac{\pi D_{n-1}}{D_c} - \frac{\pi D_n}{D_c} \quad (9)$$

However, from Fig. 7,

$$D_n = (D_{n-1} - 2D_c) \quad (10)$$

Then, from equations 9 and 10,

$$\Delta = \frac{\pi(D_{n-1} - D_{n-1} + 2D_c)}{D_c} \quad (11)$$

$$\Delta = 2\pi \quad (12)$$

Equation 12 indicates that in an ideal precision toroidal winding each layer has 2π fewer turns than the previous layer. Of course, in actual practice, Δ must be an integer number, e.g., 6 or 7, and the 2π value is achieved only by averaging over a number of layers.

Since Δ is constant and independent of the relative layer and wire diameters, equation 12, it is possible to determine the number of layers. For a winding of N turns with T_1 turns in the first layer, the arithmetic progression equation[10] is

$$N = \frac{L}{2}[2T_1 + (1-L)\Delta] \quad (13)$$

by algebra, and ignoring the extraneous root, the practical solution is

$$L = \left(\frac{T_1}{\Delta}+\frac{1}{2}\right) - \sqrt{\left(\frac{T_1}{\Delta}+\frac{1}{2}\right)^2 - \frac{2N}{\Delta}} \quad (14)$$

Thus equation 14 gives the number of layers in a precision wound toroid. By applying

Table III. Calculation Sheet Summarizing Design Calculations

AMPLISTAT CALCULATION SHEET No. Example Design

Rating 14.5 va output 52.5 ohm Load: Power 43.7 volts ±10%, 60 Cycles per Second 15%.
Signal 5.5 milliamperes in 121 ohm winding Date August 17, 1957
Core AE-7189 Anneal per O.V. Supplier I^2R Gate 6.78 D.S.
or Core MI-50012 Steel 0.002" Deltamax (*1) I^2R Control 0.0103 C.S.
#118
I^2R Total 6.79 Total Loss 7.82 N.P.

Volts	48.0	Rms		
Total Flux	18 710.	Maxwells		
Net-C.S.	0.200	in^2		
	1.29	cm^2		
Density	14 500.	gausses		
Max Milliwatts / Gram	2.68	mw/gm		
Net Wgt Iron	139.7	grams-max.		
	0.307	pounds-max.		
Max Core Loss	0.514	watts per core		
A.T.c	1.717	ampere-turns		
A.T. Feedback				
Max. Ex. Amps	1.69	milliamperes		

$\phi = B_s A_{Fe} = 14500x1.29 = 18710.$ Maxwells

$$N = \frac{E \times 10^8}{4.44 f \phi} \quad \frac{48x10^8}{4.44x57x18710} = 1014 \text{ turns}$$

	Core	Case	
OD =	1.750"	1.850"	GQV=0.2597 in^4
ID =	1.250"	1.150"	A_{Fe}=0.250 in^2
h =	1.000"	1.090"	A_w =1.0386 in^2

Note: Two cores wound as below required for each single-ended amplistat.

	GATE	CONTROL
O.C. Volts	48.0	
Line Volts (Calc.) a-c	43.7	
Load Volts (Rated) d-c	27.6	
Amperes Series	0.498	0.0055
Total Turns	1014	312
Taps	None	None
Number of Coils	One per core	One per core
Turns per Coil	1014	312
First Layer Turns	165	82
No. Layers	7	4
Traverse	360°	50°
Test V bet. L.	15.6	15.6
Conductor Bare	0.0179	0.0031
Conductor Ins.	0.0206	0.0039
Sq. in C.S.	$252x10^{-6}$	$7.5x10^{-6}$
Amp. per Sq In.	1976.	733.
Length of Turn ft.	0.3218	0.3533
Total Length ft.	326.3	110.2
Resistance	10.2/10.8/11.4	121.
IR Drop	5.67	0.6655
Lbs Copper	0.3217	0.003284
Color Code	Black-White	Black-Red
High Pot. Test.	1500	1500

ABBREVIATIONS USED IN THIS FORM

AF - after
Amp per sq in - current density-amperes/in^2
AT - ampere-turns
AT_c - dynamic coercive force-amp-turns
BF - before
CS - coil sketch number
$\frac{Cu}{Fe}$ - ratio of copper weight to iron wt.
DS - diagram sketch number
INS - insulation or insulating
MLMC - mean length of magnetic circuit
Net CS - net cross section (iron)
NP - name plate number
OC - open circuit
Sq in C.S. - (conductor) cross section-in^2
Utilization - percent of window area used by winding
V bet. L - voltage between layers

Wire Space 1.15" Dia. Gauge 25, 40 M.L.M.C. 11.99cm Volts per Turn 0.0474 Lb Copper—Total 0.325

Winding Form	Core Case	Gate&Insulation
Ins. Coil	2-$\frac{1}{2}$L-0.004"KP	2-$\frac{1}{2}$L-0.004"KP
Conductor	Double Formex (*2)	Dbl. Form. (*2)
Ins. bet. Layers	None	None
Hole BF. INS.	0.850"	0.73"
Hole AF. INS.	0.770"	0.65"

$\frac{Cu}{Fe} \frac{0.325}{0.307} = 1.028$

Utilization = 68.2%

MOD. ST-33 (8-57) Prepared by R.E.A. Checked by

(*1) - Trade Mark - Arnold Engineering Co.
(*2) - Trade Mark - General Electric Co.

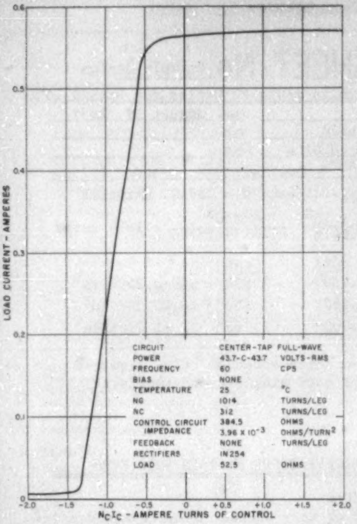


Fig. 6 (left). Measured control characteristic design

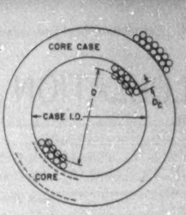


Fig. 7. Winding geometry

the appropriate winding factors to the first layer turns T_1 and to the turns difference Δ, equation 14 may be made to represent the actual performance of any winding machine. Although in practice T_1 is always slightly reduced by this winding factor, the fact that most toroidal winding machines do not build successive layers on top of the previous layer, but have a tendency to drop conductors in some of the grooves between turns of the previous layer, also makes Δ slightly reduced in practice. (Note that in a toroidal winding with layer insulation, Δ would always be greater than 2π.) For these reasons the net effect of the winding factors may become quite negligible.

Appendix II. Matching Index

As an aid in estimating $R_o/R_L=1+R_G/R_L$, one may use the similarity laws published by Storm,[3] chapter 16. These relationships are

$$\frac{P_{L,m}}{P_{L,m_1}}=\left(\frac{d}{d_1}\right)^3\left[2\left(\frac{d}{d_1}\right)^{3/2}-1\right] \tag{15}$$

from which the ratio d/d_1 may be estimated. Then

$$n=\left[2\left(\frac{d}{d_1}\right)^{3/2}-1\right]=\frac{R_L}{R_G} \tag{16}$$

and hence

$$\frac{R_o}{R_L}=1+\frac{1}{n} \tag{17}$$

Of course, these equations are based on dimension ratios, such as ID/OD, h/F, etc., which do not change with size. Furthermore, it is implied that the heat transfer geometry be independent of size. Actually in small units and large reactors mounted separately the total reactor surface is useful in cooling. On the other hand, when two medium size reactors, which comprise one magnetic amplifier, are placed on top of one another for mounting or encapsulation, only half the reactor surface may dissipate heat. Thus, in practice, the actual heat transfer geometry may change with size. In the latter case equation 17 may approach

$$\frac{R_o}{R_L}=1+\frac{1}{2n} \tag{18}$$

where n is determined from equations 15 and 16 with, as a base, a reactor where the entire surface is available for heat transfer. For instance, the base sizes are approximately 3.6 va at 60 cps and 0.14 va at 400 cps for typical toroidal designs with 55 degrees centigrade rise. (Note that the base reactor is of such size as to be both resistive and thermally limited. Thus the base reactor represents the boundary case where resistive limited and thermally limited meet.)

Appendix III. Nomenclature

A_C=area of copper allocated to control, in
A_{CC}=cross-sectional area of control conductor, in^2
A_{CG}=cross-sectional area of gate conductor, in^2
A_{Cu}=total copper area, in^2
A_{Fe}=cross-sectional area of iron, net, in^2
A_G=area of copper allocated to gate, in^2
AT_c=dynamic coercive magnetomotive force, oersteds
A_w=cross-sectional area of window, in^2
B_s=saturation flux density, gausses
d=representative dimension of a magnetic amplifier
d_1=representative dimension of base magnetic amplifier of such size as to be both resistive and thermally limited
D=mean inside diameter of a layer of winding
D_c=diameter of conductor, inches
E=effective line voltage, rms volts
f=frequency, cps
F=core stack, (OD—ID)/2 inches
g=weight per unit length of conductor material, pounds per 1,000 feet
GQV=gross quartic volume=NQV/s_f, in^4
GQV_a=gross quartic volume after modification, in^4
h=core height or tape width, inches
H_c=coercive magnetomotive force, oersteds
I_C=average control current, amperes
ID=inside diameter, inches
I_{ex}=a-c exciting current, amperes
I_G=gate-winding current, amperes
I_L=load current, d-c amperes
J=current density, amperes/in^2
k_f=appropriate form factor, no units
l_{Fe}=mean length of magnetic circuit, centimeters

Table IV. Performance Summary: Design Example

Item	Type or Magnitude	Units	Nominal Point†	Rated Points†
Circuit	center tap, 60 cps	single-phase		
Load	resistive			
Load resistance	52.5	ohms		
Supply voltage	43.7-C-43.7	volts rms		
Total control circuit, R/N^2	3.96	milliohms per turn²		
Ampere turn gain	414	at*/at		
Power gain	2,202	watts/k_f^2 watt		
Time constant	0.0542	seconds		
Dynamic power gain	40,636	$\frac{\text{w/w}}{\text{seconds}}$		
Dynamic power gain	677.27	$\frac{\text{w/w}}{\text{cycle}}$		
Power output		watts/k_f^2	10.63	16.6
Load current		amperes, average	0.450	0.563
Load voltage		volts, average	23.63	29.6
Control ampere-turns		ampere-turns from cut off	0.695	1.39
Total control power		milliwatts	1.913	7.65

* Ampere-turns.

† $I_{\text{nominal}}=0.8\,I_{\text{rated}}$; see Fig. 3 of Part I for relationships.

L = number of layers, no units
MLT = mean length of turn of wire, inches
n = matching index = R_G/R_L
N = number of turns, no units
N_C = control turns per core leg
N_G = gate turns per core leg
NQV = net quartic volume = $A_{Fe} \times A_w$ = $GQV \times s_f$, in^4
OD = outside diameter, inches
P = output power, watts d-c
$P_{L,m}$ = maximum power output, watts
P_{L,m_1} = maximum power output of base magnetic amplifier, watts
R = resistance of winding, ohms
R_C = control-winding resistance, ohms
R_G = gate winding resistance, ohms
R_L = load resistance, ohms
R_o = output circuit resistance, ohms
s_f = stacking factor, no units
t_C = thickness of control winding, inches
t_G = thickness of gate winding, inches
t_i = insulation thickness or build, inches
t_s = spool or wrapper thickness, inches
t_w = winding thickness or build, inches
T = turns per layer, no units
T_1 = turns in first layer of a winding, no units
V_L = load voltage, d-c, volts
W = weight, pounds
w/w = power gain, watts per watt
Δ = difference in T between adjacent layers, a number = 2π
π = a number = 3.1416
ρ = resistivity of conductor material, ohm-inches
P = resistance per unit length of conductor, ohms per 1,000 feet
Φ = flux, maxwells

References

1. Magnetic Amplifier Design, Part I—The Core, R. E. Anderson. *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 160–67.

2. Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers, C. L. Boyajian. *Ibid.*, vol. 75, pt. I, July 1956, pp. 307–13.

3. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

4. Thermal Characteristics of a Transformer, Final Report. *Project no. 90-686E*, Armour Research Foundation, Illinois Institute of Technology, Chicago, Ill., pp. 87 ff.

5. Temperature Distribution in Toroidal Electrical Coils of Rectangular Cross Section, T. J. Higgins. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 240, Aug. 1945, pp. 97–112.

6. Heat Transmission (book), William Henry McAdams. McGraw-Hill Book Company, Inc., New York, N. Y., third edition, 1954.

7. Simplified Design of High Temperature Transformers, Irving Romis. *Electrical Manufacturing*, New York, N. Y., vol. 56, no. 5, Nov. 1955, pp. 98–105, 334–36.

8. Electrical Coils and Conductors: Their Electrical Characteristics and Theory (book), H. B. Dwight. McGraw-Hill Book Company, Inc., 1945.

9. Toroidal Coil Developments, S. Freedman. *Radio-Electronic Engineering*, Chicago, Ill., vol. 31, Apr. 1952, pp. 3–6, 31.

10. Handbook of Mathematical Tables and Formulas, R. S. Burington. Handbook Publishers, Sandusky, Ohio, 1946.

A New Method of Frequency Error Correction in Carrier Systems

B. R. STACHIEWICZ
MEMBER AIEE

THE PROBLEM of frequency error correction in carrier systems has become more and more important in recent years. Not long ago, the use of free-running oscillators at both terminals of the system was considered satisfactory in many cases. Today an end-to-end stability of ±1 cycle is normally required.

Also, when multichannel systems grouped in a higher frequency range are used, several steps of modulation may be needed to position a channel in the proper place in the final spectrum. Since each step of modulation requires a carrier, any deviation in this carrier frequency will cause an error in the transmitted signals.

It was therefore found necessary to minimize these errors as far as possible. One of the methods presently in use consists in generating all carrier frequencies required at a terminal from a single oscillator, by harmonic or subharmonic generation of required tones. At the distant terminal, a similar arrangement is used to provide demodulating frequencies. It is evident that the frequency stability of over-all systems depends on the stability of master oscillators at the terminals. These are normally crystal oscillators enclosed in ovens to ensure constant temperature operation. Depending on the performance of the quartz crystals and ovens, adequate stability may be obtained in such systems.

However, when the system is to operate over facilities and media that may in themselves introduce changes in frequency of the signals, the foregoing method cannot correct for such errors.

There are many instances of such operation. In cable and open-wire systems, the so-called frequency-frogging repeaters are often used. Frequency frogging consists in inverting the arriving band of frequencies and transposing it into a different portion of the spectrum. In this manner equalization of losses introduced in the cable or open wire is achieved. The inversion is performed by modulation at the repeater and is shown in Fig. 1. It is apparent from this that, in a system operating through a number of frogging repeaters, a frequency error may be introduced at each repeater because of the deviation of carrier oscillator frequency from its nominal value.

It is conceivable that repeater oscillators could be enclosed in ovens to improve their frequency stability, but from a practical point of view the use of ovens at the repeaters may be impossible because of power consumption requirements and other difficulties. A good example of this is a transistorized repeater, which is powered from a distant location over the wire pairs.

The power required for the oven could possibly exceed that required for the transistor circuitry, and could also make it impossible to feed more than one or two repeaters from a distant point. The use of an oven at the repeaters is therefore not considered a satisfactory solution to this problem.

In radio communications frequency errors may also be introduced in the transmission medium. A notable example of this is the case of ground-to-air or air-to-air communication, where the receiver may be located in fast-flying aircraft.

With modern jet-propelled planes, an appreciable change in frequency may occur as the result of Doppler effect. In single side-band transmission one method of eliminating such a shift consists in transmitting the carrier frequency adjacent to the speech spectrum and using it for demodulation at the receiver. This method, however, has the disadvantage that transmission of many carriers may result in overloading the radio equipment.

Paper 58-239, recommended by the AIEE Wire Communications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted April 8, 1957; made available for printing November 29, 1957.

B. R. Stachiewicz is with the Federal Telecommunication Laboratories, Palo Alto, Calif.

The frequency-correcting circuit described has been successfully tested in the laboratory. The author wishes to acknowledge the work of O. Jorgensen, of the Stromberg-Carlson Company, who was responsible for the tests performed.

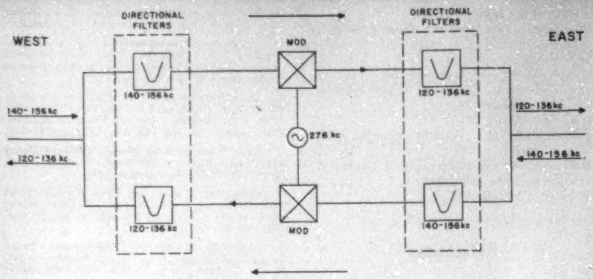

WEST
DIRECTIONAL FILTERS
MOD
DIRECTIONAL FILTERS
EAST
140-156 kc
120-136 kc
120-136 kc
140-156 kc
276 kc
120-136 kc
140-156 kc
120-136 kc
140-156 kc
MOD

Fig. 1. Frequency-frogging repeater

Group carriers, used in single side-band modulation, fall outside the transmitted band of their particular group and may interfere with other groups if transmitted.

This discussion will be confined to wire-line communications systems but it should be kept in mind that the principles are also valid for other methods of transmission.

The method of frequency error correction to be described was originally developed for a 2-wire 12-channel system. This system consists of three 4-channel groups, each group occupying 36-kc of spectrum. In this discussion only the group will be considered which transmits in the frequency band of 120 to 136 kc and receives in the band of 140 to 156 kc, or vice versa. At each repeater frequency frogging is achieved by modulation of a 276-kc carrier. The basic four channels occupy the spectrum from 12 to 28 kc. A pilot frequency located in the middle of the group (20 kc) is continuously transmitted. The simplified schematic for the complete system showing nominal frequencies is represented in Fig. 2. A 4-wire arrangement is shown in order to isolate the directions of transmission for the sake of clarity. Normally, such systems will operate on a 2-wire basis. To understand the operation of the frequency-correcting circuits to be described, it must be kept in mind that it is necessary to prevent frequency changes in the end-to-end transmission of a signal and, because the method for accomplishing this works on the principle of cancelling one error with another, carrier and pilot frequencies throughout the system will purposely contain frequency errors.

In the accompanying figures, the exact frequencies of both the signal and the pilot are given, although it is apparent that, because the pilot tone is located in the middle of the signal band, it will undergo the same changes in frequency as any other signal in the band. The reason for showing both tones is that carrier frequencies are derived from the pilot tone, and consequently a change in the frequency of this tone will result in a change of carrier frequency.

Only the common equipment portion of the terminals is shown in the figures. Frequency translation to and from the voice range is accomplished in the channel units and does not affect the discussion. Frequency errors that may occur in channel modulators are negligible, compared to those introduced at the repeaters.

It is assumed throughout this discussion that frequency frogging at the repeater is accomplished by lower side-band modulation. This is always true, because inversion of the band is required for slope correction.

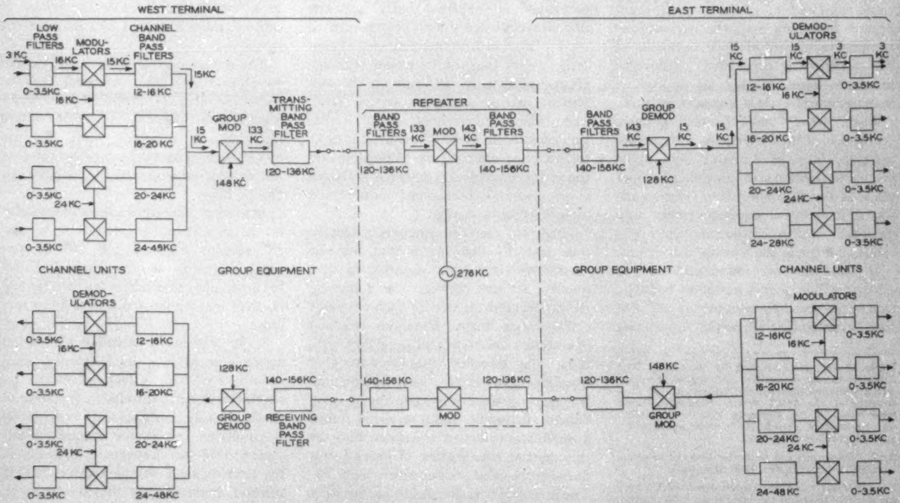

WEST TERMINAL
EAST TERMINAL
LOW PASS FILTERS
MODULATORS
CHANNEL BAND PASS FILTERS
3 KC
16 KC
15 KC
15 KC
0-3.5KC
12-16 KC
18 KC
16-20 KC
20-24KC
24 KC
24-48KC
CHANNEL UNITS
GROUP EQUIPMENT
GROUP MOD
TRANSMITTING BAND PASS FILTER
15 KC
133 KC
148 KC
120-136KC
REPEATER
BAND PASS FILTERS
133 KC
MOD
143 KC
BAND PASS FILTERS
120-136 KC
140-156KC
276KC
MOD
140-156 KC
120-136 KC
BAND PASS FILTERS
143 KC
GROUP DEMOD
15 KC
15 KC
140-156KC
128 KC
DEMODULATORS
3 KC
0-3.5KC
16 KC
24-28KC
GROUP EQUIPMENT
CHANNEL UNITS
DEMODULATORS
128 KC
140-156KC
GROUP DEMOD
RECEIVING BAND PASS FILTER
120-136KC
148 KC
GROUP MOD
MODULATORS
0-3.5KC
12-16KC
16-20 KC
20-24KC
24-48 KC

Fig. 2. Four-channel carrier system simplified block schematic

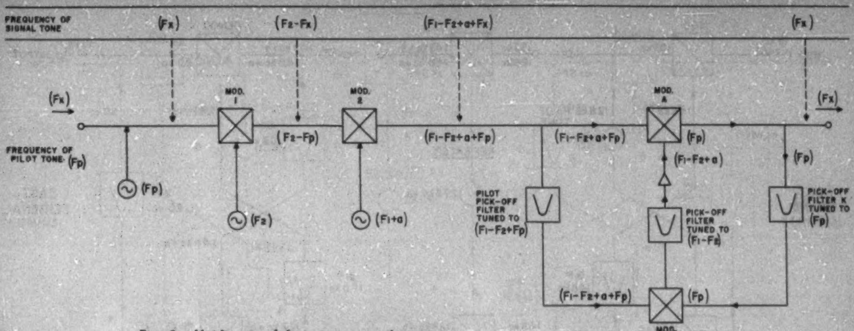

FREQUENCY OF SIGNAL TONE
(Fx)
(F2-Fx)
(F1-F2+a+Fx)
(Fx)
(Fx)
MOD. 1
MOD. 2
MOD. A
(Fx)
FREQUENCY OF PILOT TONE (Fp)
(F2-Fp)
(F1-F2+a+Fp)
(F1-F2+a+Fp)
(Fp)
(F1-F2+a)
(Fp)
(Fp)
(F2)
(F1+a)
PILOT PICK-OFF FILTER TUNED TO (F1-F2+Fp)
PICK-OFF FILTER TUNED TO (F1-F2)
PICK-OFF FILTER K TUNED TO (Fp)
(F1-F2+a+Fp)
(Fp)
MOD. B

Fig. 3. Unidirectional frequency correction

Theory

Unidirectional Frequency Correction

Assume that frequency f_p is generated by a stable oscillator and is used as a pilot in the system shown; see Fig. 3. Two stages of modulation are shown, in one of which the carrier frequency contains an error a. This one stage may be throught of as representing an unlimited number of modulators with the total frequency error introduced in all of them being equal to a. The pilot frequency therefore emerges from the modulator as $(f_1-f_2+a+f_p)$ assuming lower sideband modulation. The frequency corrective network consists of two modulators and two filters forming a "bootstrap" circuit.

The arriving pilot frequency is picked off at the input to the corrector and passed to modulator B, in which it is combined with another frequency obtained from the output of modulator A. The combination of these two frequencies in modulator B produces the required carrier frequency for modulator A. As seen from Fig. 3, the output pick-off filter K is tuned to the difference between the nominal frequency of the incoming pilot and the carrier frequency. In this manner, the carrier frequency generated from the incoming and outgoing tones contains the error a which cancels the error of the same magnitude present in the incoming signal. The output from modulator A then contain no error. Due to the bootstrap nature of this circuit, the stability of the output frequency is determined by the accuracy and stability of filter K. If filter K is tuned slightly off the f_p frequency so that f_p+a predominates, the desired signal f_x will then contain an error frequency f_x+a. However, as will be shown later, this error will cancel out when the circuit is used for 2-way trans-

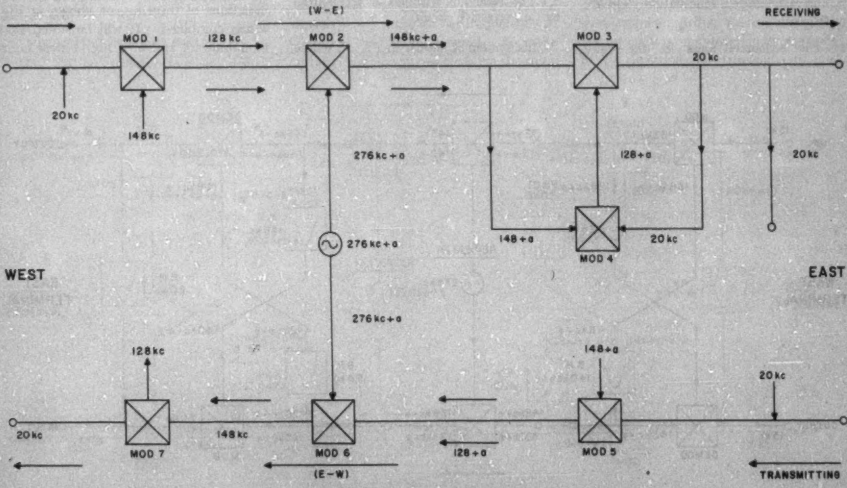

(W-E)
RECEIVING
MOD 1
128 kc
MOD 2
148kc+a
MOD 3
20 kc
20 kc
148 kc
276 kc+a
128+a
20 kc
276 kc+a
148+a
20 kc
MOD 4
WEST
EAST
276 kc+a
128 kc
148+a
20 kc
20 kc
148 kc
MOD 7
MOD 6
128+a
MOD 5
(E-W)
TRANSMITTING

Fig. 4. Principle of correction: two directions

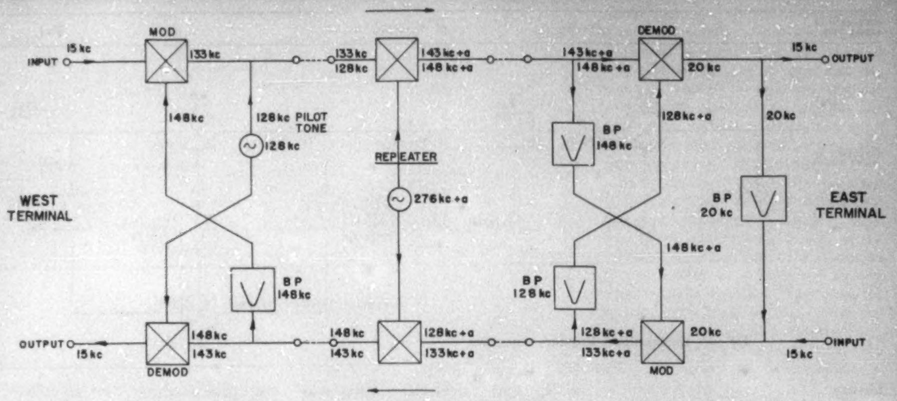

MOD
15 kc
INPUT
133 kc
133 kc
128 kc
143 kc+a
148 kc+a
DEMOD
143 kc+a
148 kc+a
20 kc
15 kc
OUTPUT
148 kc
128 kc PILOT TONE
128 kc
REPEATER
276 kc+a
BP 148 kc
128 kc+a
20 kc
WEST TERMINAL
BP 20 kc
EAST TERMINAL
148 kc+a
BP 148 kc
BP 128 kc
OUTPUT
15 kc
148 kc
143 kc
DEMOD
148 kc
143 kc
128 kc+a
133 kc+a
128 kc+a
133 kc+a
20 kc
INPUT
15 kc
MOD

Fig. 5. Combined circuitry for correction of frequency errors in both directions of transmission

mission. It is thus apparent that errors introduced in any frequency f_x transmitted over this system will be corrected in the corrector; this is also shown in Fig. 3.

Practical Applications

For the sake of clarity, this method will now be applied to the 4-channel system operating in the frequency range of 120 to 136 kc, transmitting east to west, and 140 to 156 kc receiving from the west.

The east terminal modulation steps are as follows:

1. The 4-channel units in the voice range 0 to 4 kc are first modulated to a basic subgroup of 12 to 28 kc.
2. This subgroup then modulates a 148-kc carrier to produce the output band of 120 to 136 kc. This step is represented by modulator 1 of Fig. 3.
3. The pilot frequency represented by f_p in Fig. 3 is set at 20 kc and consequently after modulation it becomes 128 kc.
4. The frogging frequency of the repeater is 276 kc and the repeater may be represented by modulator 2. An error a is assumed in this frequency.
5. The receiving terminal is represented by the corrective circuit.

This system is shown in Fig. 4, where the configuration of the W-E (west-east) circuit is identical with that of Fig. 3.

To achieve cancellation of frequency errors in the oppostie E-W (east-west) direction of transmission, the carrier frequency used in modulator 5 is purposely shifted by a cycles, which is the same error as that occurring in the repeater.

Inspection of Fig. 4 indicates that error introduced at the repeater will be canceled by the time the west terminal is reached.

In Fig. 5 the circuitry used for both directions of transmission shown in Fig. 4 was combined. It will be noted that modulators 4 and 5 of Fig. 4 were com-

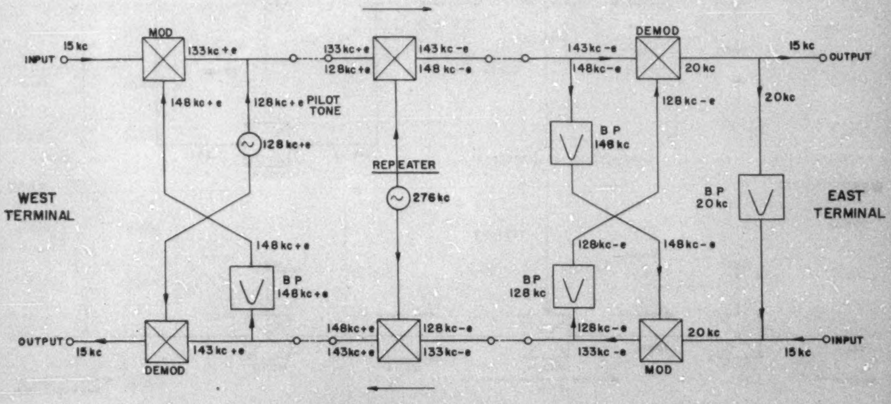

MOD
15 kc
INPUT
133 kc + e
133 kc+e
128 kc+e
143 kc -e
148 kc -e
DEMOD
143 kc -e
148 kc-e
20 kc
15 kc
OUTPUT
148 kc + e
128 kc + e PILOT TONE
128 kc+e
REPEATER
276 kc
BP 148 kc
128 kc - e
20 kc
WEST TERMINAL
BP 20 kc
EAST TERMINAL
148 kc + e
BP 148 kc+e
128 kc-e
148 kc - e
BP 128 kc
OUTPUT
15 kc
143 kc +e
DEMOD
148 kc+e
143 kc+e
128 kc -e
133 kc-e
128 kc - e
133 kc -e
20 kc
INPUT
15 kc
MOD

Fig. 6. Frequency drift of terminal oscillator

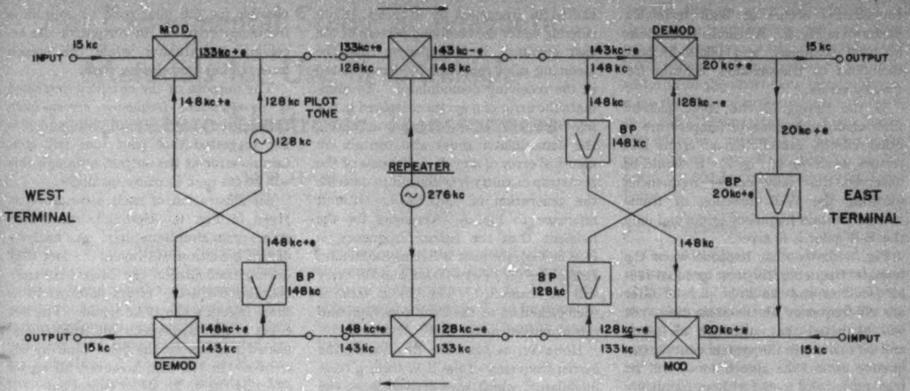

MOD
INPUT
15 kc
133kc+e
128 kc PILOT TONE
148kc+e
128 kc
WEST TERMINAL
148kc+e
BP 148kc
OUTPUT
15 kc
148kc+e
143kc
DEMOD
133kc+e
128kc
143kc-e
148 kc
REPEATER
276kc
148 kc+e
143kc
128kc-e
133kc
DEMOD
143kc-e
148kc
148kc
BP 148kc
20 kc+e
128kc-e
15 kc
OUTPUT
20kc+e
BP 20kc+e
EAST TERMINAL
148kc
BP 128kc
128kc-e
133kc
MOD
20kc+e
INPUT
15kc

Fig. 7. Detuning of the 20-kc filter

bined into one, and a 20-kc band-pass filter was introduced between the transmitting and receiving pairs at the east terminal. At the west terminal, a 128-kc oscillator was introduced to provide the receiving carrier and the outgoing pilot frequencies.

Analysis of Circuit Properties

The circuit of Fig. 5 shows very interesting error-canceling properties which will now be examined. Four conditions of interest are:

1. Only repeater frequency drifts.
2. Only 128-kc oscillator frequency drifts.
3. Only 20-kc pilot pick-off filter tuning changes.
4. All the foregoing deviations occur simultaneously.

A detailed analysis of circuit performance under these conditions is given later in this section, following a short summary.

In each case frequency correction is achieved in a demodulator where the carrier and the incoming signal contain the same error. The equation for demodulation is $(f_c+a)-(f_x+a)=f_c-f_x$, where (f_c+a) is the carrier frequency and (f_x+a) is the incoming signal. To make certain that the carrier contains the same error as the signal, the carrier frequency is derived directly from the signal.

The case of repeater frequency drift is shown in Fig. 5. A tone of 15 kc is injected at the west and east terminals. It is apparent from the figure that error cancellation results in both directions of transmission. This is true for any number of frogging repeaters. It should be noted that only the carrier frequencies at the east terminal contain frequency errors.

When only the 128-kc oscillator at the west terminal drifts, the resulting frequency error is reproduced in the opposite, E-W, direction of transmission, and

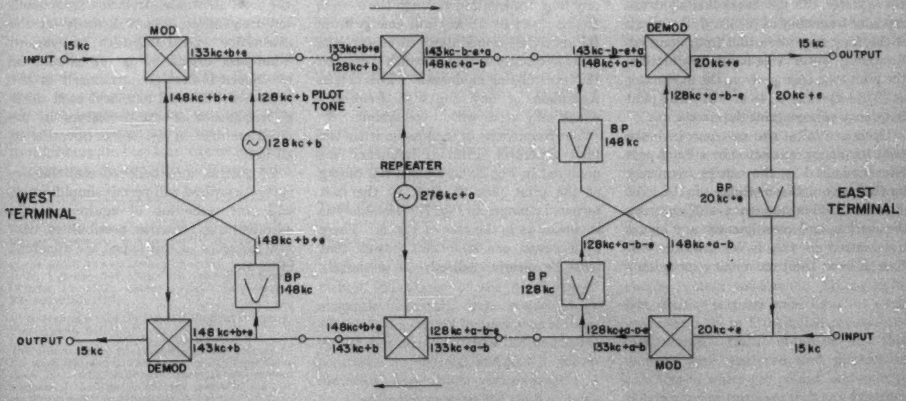

MOD
INPUT
15 kc
133 kc+b+e
148kc+b+e
128 kc+b PILOT TONE
128 kc+b
WEST TERMINAL
148kc+b+e
BP 148kc
OUTPUT
15 kc
148 kc+b+e
143kc+b
DEMOD
133kc+b+e
128kc+b
143kc-b-e+a
148kc+a-b
REPEATER
276kc+a
148kc+b+e
143 kc+b
128 kc+a-b-e
133kc+a-b
DEMOD
143 kc-b-e+a
148kc+a-b
15 kc
OUTPUT
20kc+e
128kc+a-b-e
20 kc+e
BP 148 kc
BP 20 kc+e
EAST TERMINAL
128kc+a-b-e
148kc+a-b
BP 128kc
128kc+a-b-e
133kc+a-b
MOD
20kc+e
INPUT
15 kc

Fig. 8. All errors combined

cancellation occurs at both terminals, as shown in Fig. 6. All carrier frequencies at both terminals, and pilots in both directions of transmission, contain frequency errors.

If the tuning of the 20-kc pick-off filter changes, because of temperature or other effects, cancellation of errors still results, as shown in Fig. 7. It should be observed that only carrier frequencies affecting the W-E direction of transmission contain frequency errors and only the E-W pilot is in error.

Fig. 8 shows what happens when the repeater frequency, the west terminal 128-kc oscillator and the 20-kc pick-off filter are off frequency at the same time. It will be noted that, although all pilots and all carriers in the system contain frequency errors, the speech tones will be transmitted end to end without frequency deviations. The permissible magnitude of oscillator drift is limited only by the bandwidth of the carrier pick-off filters at the terminals, which in turn depends on the permissible interference between the carrier and the adjacent speech channels.

To analyze the error-correcting properties of the circuits shown in Figs. 5 through 8, it is necessary to consider the pilot and carriers separately from the speech tones. In the figures, the frequenices of a typical speech tone are represented by 15 kc and are shown on the outside of the diagrams, i.e., above the line for the W-E direction and below the line for the E-W direction. Frequencies of the pilot tone (128 kc) and the carrier frequencies are shown inside the diagram.

The pilot tone generated by the 128-kc oscillator at the west terminal is transmitted toward the east; see Fig. 5. At the repeater, this tone is modulated by the repeater frequency of (276 kc$+a$) where a is the frequency deviation from nominal. Assuming lower side-band modulation, the pilot tone emerges from the modulator as $(276+a)-128=148$ kc$+a$. The pilot frequency now contains the error a.

Upon arrival at the east terminal, the pilot frequency is selected by a band-pass filter and used as the carrier frequency for the transmitting modulator in the E-W direction. This frequency still contains the error a and consequently any signal transmitted in the E-W direction will emerge from the transmitting modulator shifted in frequency by a cycles. Considering again the receiving portion of the east terminal, it is apparent that the incoming pilot tone will be demodulated in the receiving demodulator. To eliminate the error of a cycles contained in the pilot tone, the carrier frequency actuating the demodulator must also contain an identical error of a cycle. Because of the bootstrap circuitry it is difficult to describe the generation of this carrier without reference to Fig. 5. Assuming for the moment that the carrier frequency is (128 kc$+a$), the pilot will demodulate into $(148+a)-(128+a)=20$ kc, and the error will be cancelled. The 20-kc tone is then picked off by the band-pass filter and retransmitted in the E-W direction.

However, as has been mentioned, the carrier frequency of the E-W transmitting modulator, which was derived from the W-E pilot, contains the error a. The retransmitted pilot tone will therefore emerge from the transmitting modulator containing the error a in accordance with the following translation: $(148+a)-20=$ (128 kc$+a$).

It may immediately be seen that this frequency is the same as the one previously assumed for the W-E demodulator carrier and can therefore be used directly to actuate this demodulator, thus completing the closed loop circuit.

Considering transmission in the E-W direction, the pilot frequency containing an error a modulates the repeater carrier which also contains the same error. The lower side-band product of this modulation therefore does not contain any error. (276 kc$+a$)$-$(128 kc$+a$)$=148$ kc.

From this consideration of the pilot and carrier frequences, it may be seen that any tone transmitted through the system (represented by 15 kc) will emerge from the receiving demodulator at exactly the same frequency as that at which it entered the transmitting modulator. This is true regardless of any shifts in frequency which may occur within the system.

The frequencies of the system when the pilot generator drifts in frequency are analyzed in Fig. 6; and when the tuning of the pilot pick-off filter at the east terminal changes, in Fig. 7. The analysis is similar as in the case of Fig. 5. There is, however, one important feature that must be clearly realized: It is entirely possible for the pilot tone to contain a frequency error at the output of the receiving demodulator while the speech tone (15 kc) contains no error.

The purpose of the circuitry described is to eliminate frequency errors from transmitted speech signals only, and it is not suggested that pilot tone will contain no error at the output, although this will be the case in many instances.

An illustration of such a condition is given in Fig. 8. Here all conceivable errors occur simultaneously. An analysis of this condition will reveal the fact that during transmission the pilot frequency contains frequency errors different from those affecting the 15-kc signal. The net effect is that the speech spectrum is displaced relative to the pilot tone by an error e. In the end, however, all errors affecting the 15-kc signal cancel out.

Conclusions

The limitation of the frequency-correcting system described consists in the fact that the pilot frequency in one direction of transmission is the same as the carrier frequency required for the opposite direction. The principle may be applied, however, to a more general case by the use of additional modulators in the frequency-correcting circuits.

It should be noted that group oscillators were completely eliminated from one terminal, and that only one oscillator is required at the other terminal. Since the stability of this oscillator is not critical, no special precautions are necessary.

Oscillators in the repeaters need not be stabilized, and the number of repeaters in the system is limited only by the total allowable deviation from nominal frequencies, which depends on the selectivity and bandwidth of various band-pass filters used. It should be mentioned that errors introduced in the repeaters will tend to cancel each other if frequencies of the oscillators in the repeaters drift in the same direction in all units.

In general, it is believed that the circuitry described will permit simplification and cost reduction of equipment. In addition, many further possibilities, outside the scope of this paper, are apparent in its use.

An Experimental Study of Magnetic Materials for Use in Ultrahigh-Temperature Electronic Transformers

H. B. HARMS
ASSOCIATE MEMBER AIEE

THE VERY HIGH temperatures that are expected to be encountered in aircraft and missiles are presenting unusual problems to the designers of electronic equipment and components. Two solutions that have been suggested are the following: 1. that more cooling capacity be introduced to maintain the equipment at a safe maximum operating temperature, and 2. the use of parts and processes that are inherently temperature tolerant over a wide temperature range. If the latter approach is taken, the component designer must make a major development contribution to extend successfully the ambient temperature rating of his device from a present-day 200 C (degrees centigrade) maximum to a value which is expected to be as high as 500 C.

Through contract No. *AF33(616)-3623* the Air Force is sponsoring the development of electronic power transformers which will be capable of withstanding a 500 C ambient temperature as well as the other adverse environments to which military equipment is normally exposed. One important material used in transformers and for which the designer requires exact knowledge is the iron core. It is known to have temperature-dependent properties. The purpose of this paper is to describe the experimental work performed as one phase of the development to determine quantitatively the effect of temperature on magnetic materials. Properties pertinent to power transformers are core loss, excitation, and incremental permeability. These were measured on a grain-oriented silicon iron at several flux densities and at four different temperatures up to 550 C, to obtain basic data expected to be useful in the design of plate and filament transformers and reactors.

Materials Survey

Five general types of steels are known to retain their magnetic characteristics at temperatures as high as 600 C. These are low-carbon steel, powdered iron, silicon steel, grain-oriented silicon steel, and cobalt iron. Of these, grain-oriented silicon steel was selected for evaluation because of its well-known superior performance in normal commercial and military applications, its availability and low cost, and its long accumulated processing background. Although cobalt iron surpasses it in Curie temperature (740 C as opposed to 900 C) and in room temperature saturation flux density (15 kilogausses as opposed to 20 kilogausses), it is 20 times as costly as the grain-oriented silicon steel, and uncertain processing problems are assocated with it.

The complete designation of the silicon steel evaluated is *AISI-M6* grain-oriented $3\,^{1}/_{4}\%$ silicon steel. A 0.010-inch thickness was used because of the contemplated 400-cycle-per-second frequency application. The material was supplied in strip form with a thin film of high-temperature insulative coating on each side. It was then punched into the desired shape and given a standard 1,115 C pure dry hydrogen anneal.

Test Conditions

PROPERTIES MEASURED

Exciting current and core loss, fundamental indications of the condition of magnetic materials, were measured at 400-cycle-per-second frequency as a function of flux density at four different temperatures: 30, 250, 400, and 550 C. Measurements were made approximately 4 hours after exposure to the desired temperature. Data suitable for use in the design of reactors were also obtained at these four temperatures and at three a-c flux density values: 25, 1,000, and 5,000 gausses. Aging tests were made in addition to the short-time tests to determine the effect of exposure of the material to a 550 C temperature for 1,000 hours. The effect was measured by observing the core loss and excitation initially, at 300 hours, at 600 hours, and finally at 1,000 hours.

TEST SPECIMENS

The 0.010-inch grain-oriented strip material was punched into several sizes of *UI* and *EI* lamination shapes. These were then stacked into coils specially designed for each size. A stack height of approximately one and one-half times the leg width was used. The coils were of standard construction. The sheet and spool insulation was composed of mica paper, while a glass fiber was used for the wire insulation. The secondary was wound on first and the primary on top of it. The wire for the secondary in all samples was 0.0126-inch nickel-clad copper, while the aluminum wire of the largest diameter that could be fitted into the space available for the coil was used for the primary. In this way the coil copper losses were maintained at a negligible value. The leads were made long enough to pass through the test furnace door to eliminate junction-point problems. Six samples of each core size with their associated coils were evaluated. Representative specimens in place for the elevated temperature tests are shown in Fig. 1. The number of samples was reduced to two and to a single lamination size for the reactor design data tests. Lamination sizes for the core loss and excitation tests covered the width range for the major leg from 3/8 to $1^{1}/_{8}$ inches in 1/8-inch increments.

TEST PROCEDURE

Core loss and exciting current data were obtained with the use of the circuit in Fig. 2. The flux density was controlled with the input voltage. The induced voltage from which the flux density was calculated was measured with a calibrated rectifier volt-meter placed across the secondary leads. At each flux density setting the input current, watts, and output voltage were read. Six to eight flux density settings were used for each core over a 20- to 110-kiloline-per-square-inch range. Data obtained in this manner were analyzed and plotted as core loss in watts per pound, and excitation in ampere-turns per inch as a function of flux density and temperature.

Paper 58-209, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 25, 1957.

H. B. HARMS is with the General Electric Company, Fort Wayne, Ind.

The author wishes to acknowledge the support of the work reported in this paper by Air Force Contract No. *AF33(616)-3623* through its project engineer, J. F. Rippin, Jr.; the design of the test specimens and the performance of the tests by R. S. Russell; and the counsel and guidance of D. B. Bovey.

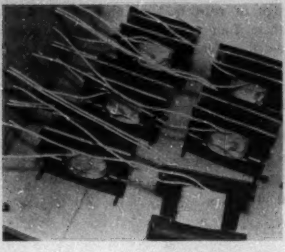


Fig. 1 (left). Group of EI punched lamination (1-inch center leg width) magnetic test specimens mounted in place for elevated temperature tests

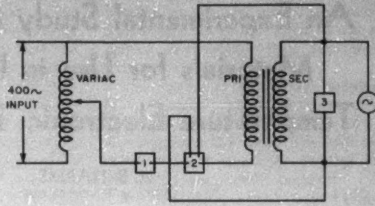


Fig. 2 (right). Test circuit for core loss and excitation tests. Instruments are: 1—thermocouple ammeter; 2—light-beam wattmeter; 3—rectifier voltmeter; and 4—oscilloscope

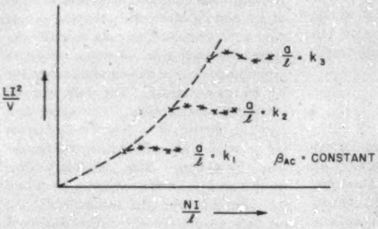


Fig. 3 (left). Typical reactor design curve

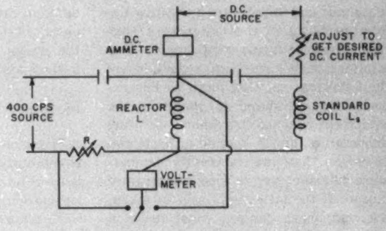


Fig. 4 (right). Test circuit for reactor design data tests

Data suitable for the design of reactors are usually plotted as LI^2/V versus NI/l[1] where

L = inductance, henrys
I = direct current flowing in the reactor circuit, amperes
V = core volume, cubic centimeters
N = number of reactor coil turns
l = mean length of magnetic circuit, centimeters

Two parameters are a-c flux density and the ratio of air gap to mean length of magnetic circuit a/l, where a = total air-gap length, in centimeters.

Data are obtained for a given air-gap ratio and a-c flux density by varying the direct current. The larger the air-gap ratio, the higher the direct current (NI/l) required to reach a maximum LI^2/V. Fig. 3 illustrates the type of raw data obtained. The envelope of the individual curves is then drawn (the dashed curve in Fig. 3) and transverse marks placed on it at the tangent points to locate specific a/l values. This latter curve is used for the design of reactors.

The circuit used to obtain the reactor design data is shown in Fig. 4. In this circuit R is adjusted to give equal voltage drops across itself and L. The alternating voltage and direct current are held constant during this operation. Under these conditions the inductance L is calculated from

$$L = \frac{L_s R}{2\pi f L_s - R}$$

where

L = unknown inductance, henrys
L_s = known air core inductance, henrys
f = a-c applied frequency
R = measured value of resistance R in Fig. 4, ohms

With L now known, the curves in Fig. 3 may be plotted.

Data Analysis

CORE LOSS AND EXCITATION DATA

The raw data were statistically analyzed with the aid of a digital computer program. By means of this program they were first converted to watts per pound, ampere-turns, and flux density. A curve of best fit for a given core size and property was then determined by trying successively higher order polynomial equations until the 5% point of the F distribution test could be satisfied. Flux density was used as the independent

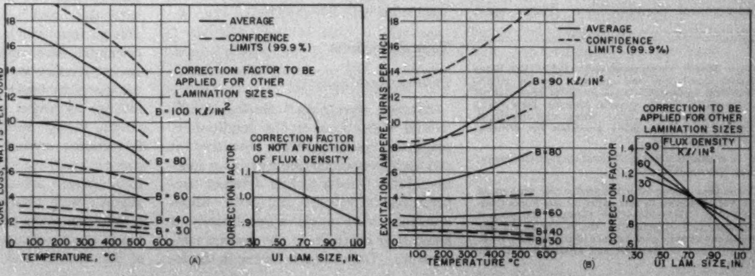


Fig. 5. Core loss and excitation as functions of temperature and flux density for UI shaped laminations. Laminations were stacked in inter-leaved alternate groups - of - four arrangement

A—Core loss
B—Excitation

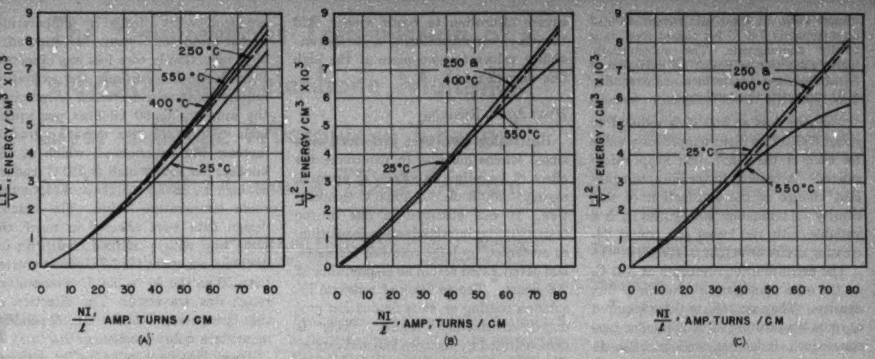


Fig. 6. Reactor design curves for various a-c flux densities: A, B, and C: 25, 1,000, and 5,000 gausses respectively

variable. Confidence limits at 99.9% were finally determined, and this and the average value were printed out at flux density points from 30 through 110 kilolines per square inch at intervals of 10 kilolines per square inch.

The core loss and excitation data as read from the computer program are plotted in Fig. 5 as a function of temperature with flux density as an adjustable parameter for particular *UI* lamination-shaped leg widths. The 3/4-inch width selected fitted approximately midway in the size range tested. Further analysis showed that data from the other lamination sizes were similar to these, although displaced in an orderly pattern depending on size and, in the case of the excitation, on the flux density as well. Simple correction factors therefore make the one set of data applicable for all the other sizes.

These correction factors, described in the foregoing, have been included in the curves of Fig. 5.

Reactor Design Data

These data were plotted according to the scheme illustrated in Fig. 3 for each of the two samples. The average envelope curve was then drawn for each a-c flux density and temperature condition. The analyzed data are plotted in Fig. 6(A) through (C). Each graph is for a different a-c flux density, while temperature is used for the parameter. The transverse gap ratio marks as used in Fig. 3 have been omitted, and replaced by a new curve, Fig. 7, of a/l versus NI/l which is a more accurate and a simpler way of determining the air gap required for a given direct current bias.

Data Interpretation

Core Loss and Excitation

The data reflect three major changes[1] that occur in magnetic steel when the temperature is increased. These are:

1. The electrical resistivity increases, giving rise to lower eddy currents and therefore a lower core loss.
2. Within the saturation limits of the material the permeability increases. The lower excitation values at an induction of approximately 50 kilolines per square inch and smaller, shown in Fig. 5(B), are a direct outcome.
3. The flux density value at which the material becomes saturated gradually de-

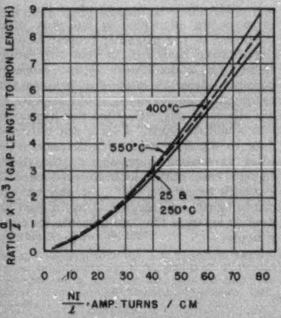


Fig. 7. Curves to determine gap length for reactor designs. Data applicable for design purposes for any a-c flux density up to 5,000 gausses

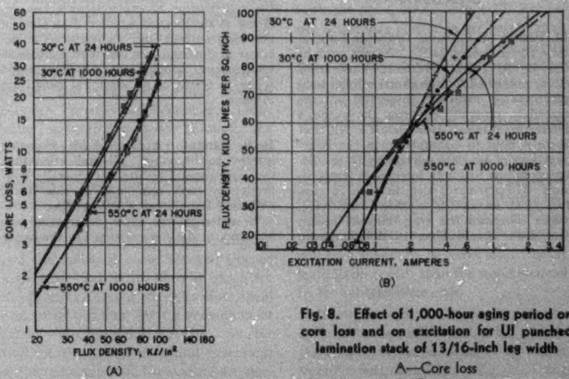


Fig. 8. Effect of 1,000-hour aging period on core loss and on excitation for UI punched lamination stack of 13/16-inch leg width

A—Core loss
B—Excitation

creases until, at the Curie temperature of approximately 740 C for silicon steel, it is totally nonmagnetic. The substantially higher excitation values at an induction of 50 kilolines per square inch and over, Fig. 5(B) emphasize this tendency.

The core loss at 550 C is reduced to about 60% of its room temperature value for each of the lamination shapes considered. According to the data, this percent change is not a function of flux density. Lamination size, however, is a variable with the losses per pound decreasing as size or weight is increased.

The excitation requirements at 550 C, on the other hand, are a function of flux density. They are 75% of those required at room temperature for a 30-kiloline-per-square-inch induction, and increase to approximately 200% at 100 kilolines per square inch. The flux density for which there is no change throughout the temperature range is estimated to be between 50 and 60 kilolines per square inch. Lamination size is a variable with the excitation decreasing as size or weight is increased.

REACTOR DESIGN DATA

Temperature and induction effects on the material are reflected in these data in the same manner as was observed with the core loss and excitation data. For a given temperature the permeability of the iron increases with flux density to a maximum until saturation starts setting in. Thereafter it slowly drops off to unity. Superimposed on this is the effect of temperature, which produces higher permeabilities at the low flux densities but causes saturation to occur sooner, and therefore, correspondingly lower permeability values to take place at the high flux densities.

HEAT AGING RESULTS

Heat aging tests were performed on a single lamination size to determine the effect of long-time exposure, of approximately 1,000 hours, to a 550 C temperature. It was observed in one of the tests that the interlamination insulation, as evidenced by large core loss increases, had deteriorated within an elapsed time of 200 hours. The addition of a second insulation coating on each lamination produced satisfactory results. This is demonstrated by the core loss and excitation data plotted in Fig. 8 at 30 and 550 C temperatures and at 24 and 1,000 hours' time exposure. The core loss for these particular conditions was essentially unaffected by the aging period. The excitation current, however, experienced small permanent changes amounting to as much as 15% at 550 C and 30% at room temperature.

Heat aging tests have indicated the very critical role that the interlamination insulation plays. The separation must be a positive one and free from any voids that will allow oxidation to progress between adjacent laminations.

Conclusions

The experimental work reported in this paper indicates that the magnetic properties of grain-oriented $3^{1}/_{4}$% silicon steel are significantly affected by temperature. In the 30 to 550 C range investigated, a 60% reduction in core loss was observed. For identical conditions the excitation requirements were decreased or remained the same up to 60 kilolines per square inch. Above this flux density a gradual increase was experienced, which at 100 kilolines per square inch at 550 C reached as high as 200% of the corresponding room temperature value. The reactor design data were affected in much the same way with a marked change in inductance occurring for NI/l values greater than 50 as the 30 to 550 C temperature range was traversed. The direction of this change was found to be dependent upon the a-c flux density.

These data have value for the designer who is confronted with the problem of designing small electronic power transformers and inductors for operation in a wide range of ambient temperatures. Possible revisions in his design practice that are a direct outcome of this knowledge include the use of lower operating flux densities, adjustment of the temperature rise to take into account the change and redistribution of the losses in both the core and coil, and a compromise on the gap setting in the case of an inductor application.

References

1. DESIGN OF REACTANCES AND TRANSFORMERS WHICH CARRY DIRECT CURRENT, C. R. Hanna. *AIEE Transactions*, vol. 46, 1927, pp. 155–60.

2. FERROMAGNETISM (book), R. M. Bozorth. D. Van Nostrand Company, Inc., Princeton, N. J., 1951, pp. 713–19.

◆

Discussion

M. Pasnak (U.S. Naval Ordnance Laboratory, Silver Spring, Md.): Current work at the U. S. Naval Ordnance Laboratory (to be published) on the effects of high temperatures on magnetic properties of core materials, generally confirms the author's interesting results. However, it is felt that there is need to elaborate upon the interpretation of his data pertaining to the decrease in core loss with increase in temperature. He attributes the decrease in core loss to an increase in electrical resistivity, which decreases the eddy current losses. Our data, obtained on 12 different core materials tested up to 800 C, indicate that the contribution of the hysteresis loss decrease is significantly greater than that of the eddy current loss. Therefore the decrease in the total core loss with increase in temperature should not be interpreted solely in terms of an electrical resistivity increase.

H. B. Harms: I appreciate Mr. Pasnak's interest in the paper. His indication that the experimental results are generally confirmed by tests conducted at the Naval Ordnance Laboratory is encouraging.

No attempt was made during the performance of the tests reported in the paper to separate the core loss into its hysteresis and eddy-current components. The test program was designed to find out what happened to the magnetic properties on a quantitative basis and to view the causes simply on a qualitative basis. Since it becomes easier to magnetize iron as the temperature increases it would appear, as the discusser suggests, that the hysteresis loss would decrease. I have not obtained data to say how much this decrease is or what its content is relative to the eddy-current loss, and I look forward to Mr. Pasnak's report on his work and on the measurement and analytical techniques he used to separate the core loss into its components.

The Simultaneous Transmission of Television and Telephone Multiplex Over a Single Microwave Channel on the Trans-Canada TD-2 System

H. E. CURTIS
NONMEMBER AIEE

U. C. P. STRAHLENDORF
NONMEMBER AIEE

A. J. WADE
NONMEMBER AIEE

THIS YEAR, a *TD-2* radio relay system[1] will span Canada from Halifax to Victoria. Certain sections of the network are designed to handle television and telephone circuits simultaneously on the same broad-band channel. It is the purpose of this paper to describe the transmission and terminal-equipment aspects which make this new technique possible and economically attractive.

The idea of stacking or "double-decking," as this technique is termed, was prompted by a demand for a limited number of long-haul telephone circuits coincident with a requirement to provide nationwide television service. To test the feasibility of stacking and to accumulate operating data, a successful field experiment was conducted by the Bell Telephone Laboratories early in 1954 on a 1,133-mile *TD-2* route with the use of standard telephone and television terminal equipment. In this experiment, two conventional high-pass–low-pass filters were used to combine and separate the signals at the *TD-2* terminals. Field results indicated that two or three telephone supergroups,[2] of 60 channels each, could be stacked in the 6- to 7-mc (megacycle) region above the television signal. While the tests and study indicated that NTSC (National Television Systems Committee) color programs could probably be satisfactorily transmitted over a stacked system, the final decision to proceed with the stacked installation was based on the assumption that the system would be unstacked before color programs were transmitted.

Fortunately, a *TD-2* channel and its associated FM (frequency modulation) transmit and receive equipment can be employed without modification. The terminal video circuit arrangements, however, require considerable engineering in order that these two distinctly different types of service may be properly combined and transmitted simultaneously. The video installation arranged for stacking is a specially engineered job at most radio terminals. Variations arise due to: 1. different telephone traffic and television broadcasters requirements, 2. the existence of three types of combining filters, and 3. transmission considerations to maintain high-quality circuit performance that long-haul communication service demands.

Fig. 1 illustrates the geographical position of the Trans-Canada *TD-2* route with the stacked sections shown in heavy lines. On September 30, 1956, television was inaugurated on the Toronto–North Bay–Winnipeg sections. The progressive addition of telephone circuits began in December. One supergroup is provided for Toronto–North Bay telephone service and two supergroups operate between Toronto and Winnipeg. These latter supergroups pass through the North Bay telephone terminal at intermediate carrier frequencies. The *TD-2* channels between Toronto and North Bay are thus loaded with 60 more stacked telephone circuits than elsewhere on the system. The westward route to Calgary and the eastward route between Quebec and Saint John, N.B., both of which are in operation, handle one television signal each way together with a maximum of 120 2-way telephone circuits. As the traffic load approaches its limit on sections of the radio network, the stacked arrangement will be discontinued and each service will be carried on separate radio channels.

The transmission considerations, terminal circuits and tests, associated with the stacked arrangement of television and telephone circuits, will be described in detail, followed by a discussion of a few of the important transmission problems that confronted the engineer when the first stacked route went into service in the autumn of 1956.

Transmission Considerations

FREQUENCY ALLOCATION

The *TD-2* System[1] is designed to transmit a base-band signal extending from virtually zero frequency to approximately 10 mc. In the standard system the signal is either a television signal whose significant energy extends from the bottom of the band to about 4 mc, or a block of 480 to 600 telephone channels arranged in frequency division multiplex and extending from about 64 kc to about 2 to 3 mc.

In the stacked arrangement described in this paper both a television signal and a limited number of telephone circuits are combined for transmission simultaneously over a single broad-band radio channel. The frequency allocation plan is shown in Fig. 2. As before, the television signal extends to slightly over 4 mc. The telephone channels extend from 6,012 to 6,748 kc and provide a maximum of 180 channels. These are derived from channelizing equipment developed for the *L-3* coaxial carrier system.[3]

RADIO CHANNEL NOISE CONSIDERATIONS

The typical noise spectrum of a *TD-2* channel is shown in Fig. 3. This represents the noise per cycle in a *TD-2* channel, arbitrarily plotted in decibels relative to the noise at 5 mc. This noise largely arises in the converter and preamplifier associated with each *TD-2* repeater, the absolute magnitude of the noise depending upon the number of repeaters in the circuit.

It will be noted that over about 500 kc the noise power per cycle increases 6 db (decibels) per octave of frequency. With the same FM deviation per telephone channel, the double-decked channel may therefore be expected to be at least 10 db noisier than the worst channel in the all-telephone case.

However, if pure FM were used, the noise in the worst channel in the all-telephone case would be some 15 db higher than the best channel; see Fig. 3. Using pre-emphasis, the deviation of the top

---

Paper 58-183, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted July 8, 1957; made available for printing November 29, 1957.

H. E. CURTIS is with the Bell Telephone Laboratories, Inc., New York, N. Y., and U. C. P. STRAHLENDORF and A. J. WADE are with the Bell Telephone Company of Canada, Montreal, Que., Canada.

The achievement described in the paper was made possible through the close co-operation of all personnel concerned in the Bell Telephone Laboratories, the Northern Electric Company, the Western Electric Company, the American Telephone and Telegraph Company, and the Bell Telephone Company of Canada.

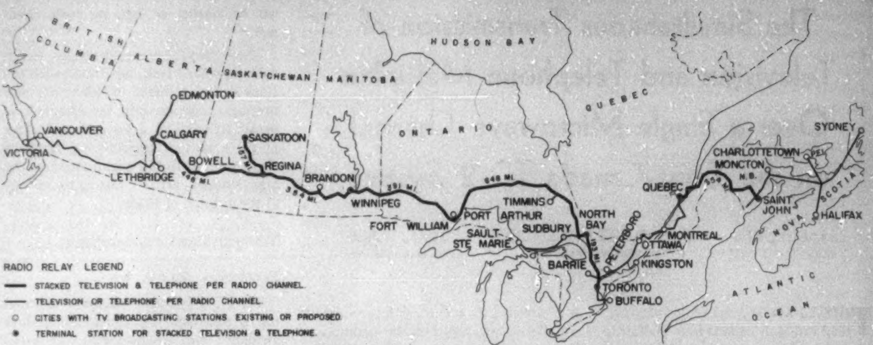

BRITISH COLUMBIA
ALBERTA
SASKATCHEWAN
MANITOBA
HUDSON BAY
ONTARIO
QUEBEC
EDMONTON
VANCOUVER
VICTORIA
CALGARY
LETHBRIDGE
BOWELL
SASKATOON
REGINA
BRANDON
WINNIPEG
FORT WILLIAM
PORT ARTHUR
TIMMINS
SAULT-STE MARIE
SUDBURY
NORTH BAY
BARRIE
PETERBORO
TORONTO
BUFFALO
KINGSTON
OTTAWA
MONTREAL
QUEBEC
SAINT JOHN
N.B.
MONCTON
CHARLOTTETOWN
P.E.I.
SYDNEY
HALIFAX
NOVA SCOTIA
ATLANTIC OCEAN
RADIO RELAY LEGEND
STACKED TELEVISION & TELEPHONE PER RADIO CHANNEL
TELEVISION OR TELEPHONE PER RADIO CHANNEL
CITIES WITH TV BROADCASTING STATIONS EXISTING OR PROPOSED
TERMINAL STATION FOR STACKED TELEVISION & TELEPHONE

Fig. 1. Trans-Canada TD-2 radio system television and telephone network routes

channel may be increased and the deviation of the bottom channel decreased, thereby obtaining an improvement in noise in the worst channel of approximately 3 to 4 db. Since the inherent noise on the top double-decked telephone channel is only 1 db greater than the noise on the bottom channel, pre-emphasis networks in the wire line entrance link are of negligible value and are therefore not installed. With the same deviation per channel, the total fluctuation noise in the double-decked channels may not be 10 db higher but actually about 13 to 14 db higher than in the all-telephone case. The field test results showed that this noise impairment may be reduced to 7 to 8 db by increasing the normal level of the *L* carrier 6 db and decreasing the normal television level 3 db at the FM transmitter input. Adverse intermodulation effects, as the result of these level changes, were not evident during the test.

Intermodulation Considerations

The frequency-modulated carrier may be distorted in amplitude and phase to a small degree by any amplitude and phase distortion in the IF (intermediate frequency) and RF (radio frequency) portions of the system and this gives rise to intermodulation products in the detected base band and signal. These products may originate within the television signal or the telephone channels, or from the combination of both signals. Intermodulation products produced by the television signal alone may degrade the picture itself and may also introduce spurious energy at the second harmonic of the color carrier. Products involving the telephone channels alone may fall in the band occupied by the television signal or may fall back into the telephone channels themselves. Other products involve both the television signal and telephone channels simultaneously.

A monochrome television signal produced by scanning is composed of concentrations of energy located in frequency regions harmonically related to the line and field-scanning frequencies. Experience has indicated that with the normal degree of phase and amplitude equalization achieved in the *TD-2* system, the intermodulation distortion is unimportant.

The NTSC color signal has, in addition to the frequency components just mentioned, a strong component of energy at approximately 3.58 mc with concentrations on either side of it. This energy is spaced at intervals of the line-scanning frequency of approximately 15.75 kc; see Fig. 2. Amplitude and phase distortion of the FM wave give rise to a distortion spectrum of this energy centered at 7.2 mc, with rapid fall-off on either side. The magnitude of this distortion spectrum depends largely on the residual linear envelope delay distortion in the RF and IF portions of the system remaining after normal phase equalization.

Sufficient frequency space between the top of the television signal and the bottom of the band allotted to the telephone channels must be allowed in order to combine and separate them without excessive phase distortion in the television signal. Furthermore, the top end of the telephone band must be somewhat under 7.2 mc so that the energy in the distortion spectrum of the NTSC color signal, if transmitted, will not interfere significantly with the telephone channels. For these reasons, the telephone channels have been restricted to a frequency bandwidth of 736 kc lying between 6,012 and 6,748 kc.

Second-Order Distortion Products

Amplitude and phase distortion of the FM wave may produce all orders of distortion of which the second order has been found to be the most significant in the present application. In particular,

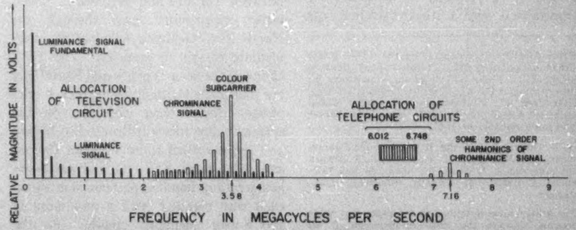


Fig. 2. Frequency composition of NTSC color television signal with telephone circuits arranged for stacking on the same TD-2 channel. One out of every ten frequency components is shown

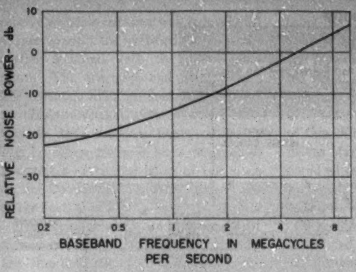
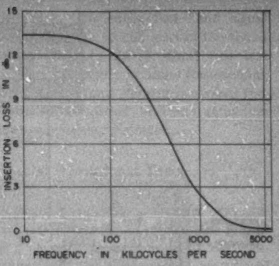


Fig. 3 (left). Relative noise power per cycle at FM receiver

Fig. 4 (right). Loss frequency characteristic of television pre-emphasis network



second-order difference products involving the telephone channels themselves, will fall into the 4.2-mc bandwidth occupied by the television signal. Since the telephone channels are spaced at 4-kc intervals, this interference spectrum will begin at a base-band frequency of 4 kc, where the number of possible difference products is a maximum, and will extend to a frequency of 736 kc equal to the band covered by the telephone channels. These products may appear as interference in the received television picture. This interference can be reduced by dropping the level of the telephone channels or by raising the level of the television signal at the FM transmitter. Lowering the telephone level is more effective.

In addition to the intermodulation products just discussed, there will also be products involving both the television signal and the telephone channels simultaneously. The most critical of these is a second-order sum and difference product involving the strong component of the television signal at approximately 15.75 kc and any particular telephone channel frequency. Since the telephone channels are spaced at 4-kc intervals, these products will appear in the telephone channel located 15.75 kc away from the channel originating the product, and will appear as intelligible although distorted crosstalk. Similarly, other products involving harmonics of the line scan frequency of 15.75 kc will be generated.

OPERATING LEVELS AT FM TRANSMITTER

The relative operating levels of the combined television and telephone channel signals at the FM transmitter input are based on a compromise of the interacting effects discussed in the foregoing. Thus, if the telephone channel level is dropped, interference into the television signal is reduced. However, while the intelligible telephone crosstalk remains unchanged, the amount of fluctuation noise in the telephone channels increases. If, instead, the television signal level is lowered, interference into the telephone channels is reduced, but distortion from the telephone channels into the television channels, as well as fluctuation noise, is increased.

The inherently lower noise at the low end of the video band achieved by the use of FM has permitted the use of a substantial amount of pre-emphasis of the television signal, as shown in Fig. 4. This reduces the relative deviation of the low-frequency components of the television signal by about 14 db and affords a corresponding reduction in the intelligible crosstalk in the telephone channel without appreciable penalty to the television signal.

The optimum relative operating levels of the two signals represents a balance between these various factors. Satisfactory intermodulation performance, however, imposes tolerances on the amplitude and phase distortion in the IF and RF band which are substantially more restrictive than when the signal is all-telephone or television only.

COMBINING FILTERS

The art of juxtaposing isolated blocks of information in a given frequency band developed contemporaneously with the conception of carrier telephony, but a new technique is involved in transmission via radio of an NTSC television signal simultaneously with a limited but relatively large number of long-haul circuits.

A combining filter, designed specifically for the purpose of combining and separating the two bands at transmitting and receiving terminals, accomplishes this operation by means of its low- and high-pass filter sections. The low-pass television circuit of the combining filter must be equalized for constant loss and constant phase delay at all frequencies up to 4.5 mc. Also, high-frequency ringing in the television path should be minimized. This requires the delay equalization to extend well into the cutoff region of the low-pass filter; see Fig. 5.

In the initial stages of the development, a preliminary combining filter was designed. Subsequent testing indicated that not more than eight of these filters could be used in tandem. An improved filter design was therefore developed which gave less distortion in the through circuit. This new branching filter performed the same function as the original filter. A combination of these two filters in tandem is required in all video receiving circuits at terminal points to suppress undesirable high-frequency delay distortion; see the receive circuit arrangement in Fig. 6.

The components and the method of assembling the filters are the same as those used in filters for the *L3* system television terminals.[8] The filters provide a communication system which gives television transmission through a number of stacked sections in tandem comparable in quality with that of a standard *TD-2* system.

Terminal Equipment

A fundamental requirement in the design of telephone and television transmission systems is that any modification

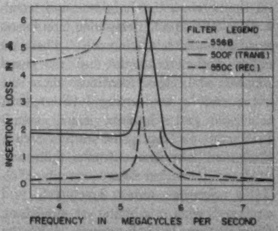


Fig. 5. Insertion losses of combining filters

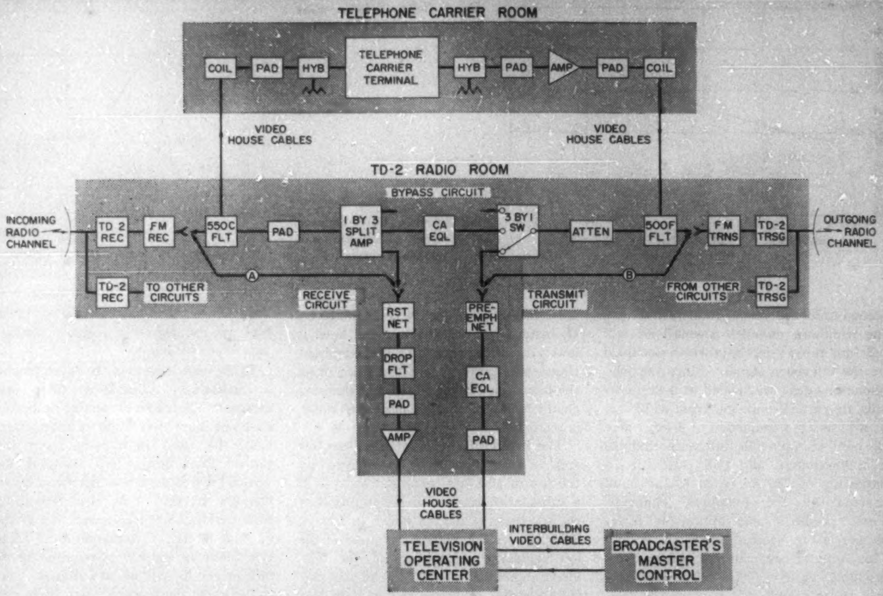

TELEPHONE CARRIER ROOM
COIL
PAD
HYB
TELEPHONE CARRIER TERMINAL
HYB
PAD
AMP
PAD
COIL
VIDEO HOUSE CABLES
VIDEO HOUSE CABLES
TD-2 RADIO ROOM
BYPASS CIRCUIT
INCOMING RADIO CHANNEL
TD 2 REC
FM REC
550C FLT
PAD
1 BY 3 SPLIT AMP
CA EQL
3 BY 1 SW
ATTEN
500F FLT
FM TRNS
TD-2 TRSG
OUTGOING RADIO CHANNEL
A
B
TD-2 REC
TO OTHER CIRCUITS
RECEIVE CIRCUIT
TRANSMIT CIRCUIT
FROM OTHER CIRCUITS
TD-2 TRSG
RST NET
PRE-EMPH NET
DROP FLT
CA EQL
PAD
AMP
PAD
VIDEO HOUSE CABLES
INTERBUILDING VIDEO CABLES
TELEVISION OPERATING CENTER
BROADCASTER'S MASTER CONTROL

Fig. 6. Video interconnecting circuits at TD-2 radio terminal

or new development must be capable of operating with previously installed systems. This means that provision must be made to interconnect new systems into the pattern of the existing plan. The establishment of a *TD-2* radio relay system requires interconnections with telephone terminal equipment and a television operating center[4] (TOC) at terminal stations. This is accomplished with "wire line entrance links" and "video interconnecting circuits" respectively.

Standard wire line entrance links[5] and video interconnecting circuits are installed at a number of *TD-2* terminals across Canada. They appear at those terminals on the network that originate and terminate television and telephone circuits on separate radio relay channels. There are eight major terminals and an equal number of bridged locations where the video circuits are arranged to transmit television and telephone signals simultaneously on the same radio circuit. Simplified wire line entrance links and modified television interconnecting *TD*-TOC circuits were developed and equipped with combining or branching filters at these latter locations. With the exception of these filters, the video interconnecting circuits require no basic developmental work but considerable rearrangement including additional standard components.

Video Interconnecting Circuits

One or more of three types of video interconnecting circuits are installed at all *TD-2* radio terminals. The TOC-to-*TD-2* radio interconnecting circuit transmits television signals from a television operating center to an FM transmitter in the nearby *TD-2* radio room. The *TD-2*-to-TOC video circuit accepts the television from the FM receiver and transmits the signal to the local TOC. The video bypass circuit, which is a new development, essentially interconnects an FM receiver with one or two FM transmitters in the same *TD-2* radio room. It also provides interconnections with both foregoing interconnecting circuits.

Video bypass circuits are installed at *TD-2* terminals, arranged for stacked television and telephone circuits on the backbone radio route. The geographical locations are Saint John, Quebec, Toronto, North Bay, Winnipeg, Regina, and Calgary. The bypass circuit avoids the necessity of routing via the TOC those incoming television signals destined for distant points. Repeated de-emphasis and pre-emphasis, Fig. 4, of the television signal is avoided and at a number of terminals this circuit permits the replacement of the combining filter with branching filters in the through circuit. The number of combining filters in tandem is therefore held to a minimum along the microwave route. The quality and reliability of picture transmission are hence retained with the bypass circuit.

A typical video bypass circuit is illustrated in Fig. 6. The equipment arrangement at some locations is more complex than at others and may include three or four different circuit variations. In contract, other *TD-2* terminals, mostly those points bridged off the main route, are equipped with simple received and transmit interconnecting circuits. When the various sections are unstacked, the interconnecting circuits will be restored with the bypass circuit and dropping filter removed.

Wire Line Entrance Links

In contrast to the numerous video interconnecting circuits for stacked *TD-2* channels, only one type of transmit and receive wire line entrance link is developed to interconnect the modified *L3* telephone carrier terminal with the *TD-2* radio room. The transmit and receive wire line entrance links are installed at *TD-2* terminals on stacked sections of the Trans-Canada route shown in heavy outline in Fig. 1. These entrance links, designed to transmit a restricted band of telephone carrier frequencies, are illustrated in the upper portion of Fig. 6.

The relatively large number of *L* terminals and TOC's along the route, and the impracticability of combining and separating these two services at IF, have produced a variety of terminal interconnecting circuits in an effort to maintain the over-all transmission objectives. While some additional units have been required as a result of this variety they will all find ready use in expanding TOC's when the stacked arrangements have served their purpose.

Initial Transmission Tests

All microwave channels of a *TD-2* system are normally subjected to a number of transmission tests prior to service. These include for example, measurements of video response, base-band noise, 2-tone distortion, and differential phase and gain. Tests and adjustments are also made on the FM transmitting and receiving terminals to ensure that they do not produce any undue transmission degradation of the signal.

Phase correction is normally provided at the 70-mc IF frequency in order to achieve adequate intermodulation performance when the system is used for telephone transmission. However, the application of stacking to a *TD-2* radio channel imposes stricter requirements on phase distortion on the system, if the performance of the video and telephone circuits is to be satisfactory. Thus certain special preservice tests were made on the stacked system. First, the over-all phase characteristic of the system was examined most critically in order to identify and eliminate, as far as feasible and necessary, the various sources of envelope delay distortion; second, the optimum input levels to the FM transmitter were determined experimentally for best performance of both the television and telephone channels. These tests are discussed more fully in the following.

ENVELOPE DELAY DISTORTION

As the first step in minimizing the envelope delay distortion, a visual inspection was made of the curve of envelope delay distortion in the microwave system, using the visual delay set, which is standard test equipment for the *TD-2* system, and is arranged to swing the RF carrier ±7 mc at a rate of 100 cycles per second, and to superimpose upon the signal a 277.78-kc "flutter" signal. The phase of this flutter signal as the FM carrier moves throughout the IF band, is compared with its average phase. This comparison is displayed upon an oscilloscope at the receiving end.

When the visual delay set was used on the looped Toronto–North Bay–Toronto circuit, it was evident that the circuit required envelope delay distortion slope equalization. Further, while envelope delay distortion ripple was observed on all channels, it was found to be especially severe on one particular channel. Specifically, this distortion consisted of ten ripples across the IF band with a peak-to-peak amplitude of approximately 100 millimicroseconds. Calculations showed that this was well in excess of the design objective, and it was concluded that some fault in the form of an impedance mismatch existed on the system. By looping back the signal from the visual delay set at each auxiliary repeater station and noting the change in distortion ripple, a fault was located at the fifth repeater station west of Toronto. This proved to be a mismatch at the antenna caused by an improper connection of the flexible wave guide.

The visual delay set permitted phase equalization of a looped circuit on a visual basis and thus increased the accuracy with which phase equalizers could be placed in each channel.

TWO-TONE TEST

It was apparent that the major source of intermodulation in the telephone channels was second-order distortion involving any particular telephone channel and the strong component of energy in the television signal at 15.75 kc. Therefore a 2-tone test was used to evaluate, on an out-of-service basis, the amount of speech intermodulation to be expected for various amounts of delay distortion. This test consisted of simulating the *L* carrier and television frequencies by applying a 6.5-mc sine wave and a 15.75-kc sine wave into their respective inputs of the filter. By noting their sum or difference products on a narrow-band transmission measuring set at the receiving end of the *TD-2* system, it was possible to observe and evaluate the changes in the product level as phase equalizers were inserted in the system.

While, strictly speaking, in FM the amplitude of any modulation product may depend on the entire base-band signal rather than the specific part of the signal which produces the product, this test appears sufficiently accurate for assessing the potential improvement in system performance associated with any reduction in envelope delay distortion.

Special Operational Tests

Regular television service was inaugurated between Toronto and Winnipeg about 2 months prior to the commencement of the special operational tests. These tests were carried out to determine final optimum operating levels, to measure intermodulation between the television and carrier, and to measure noise conditions.

Initial test loading of the system was arranged during non-TV periods by using an Indian-head monoscope pattern as the television component and by bridging three supergroups (180 voice channels) off the Toronto-Montreal *TD-2* system for use as the *L3* carrier component. A continuous recording of this latter loading was kept in order to determine the periods of heaviest telephone usage. It was thus possible to test the system under peak-load conditions, which could be verified by reference to the recording.

The first circuit, arranged for testing under the afore-mentioned conditions, consisted of a Toronto–North Bay–Toronto loop, covering a total distance of approximately 400 miles. By using a looped circuit, it was possible to compare the quality of the outgoing and incoming television picture and to obtain noise readings on the individual voice channels of the *L3* carrier system. During these tests, observations were also made at North Bay.

As was anticipated, the relative deviations for the television and telephone channels for optimum deviation had to be determined experimentally under actual operating conditions. When the television signal produced too high a deviation, the cross-modulation in the telephone channels was unacceptable. When it was too low, the television signal was unacceptable. An optimum balance between the two was reached and then, by an intensive program of reducing the envelope delay distortion, as mentioned previously, the desired performance standards for both television and telephone channels were achieved.

The Toronto-Winnipeg system is approaching its stacked capacity and engineering is well under way for the provision of a third radio channel. This was necessary because present traffic requirements for Trans-Canada telephone circuits have exceeded the original forecast.

Experience gained to date shows that stacking of television and telephone circuits is entirely feasible and is economically attractive where the telephone circuit requirements are limited and system lengths range from about 200 to 1,000 miles. The application of stacking has proved a worth-while contribution economically, and the transmission objectives which were defined as acceptable in the initial planning stages have been met.

References

1. The TD-2 Microwave Radio Relay System, A. A. Roetken, K. D. Smith, R. W. Friis. *Bell System Technical Journal*, New York, N. Y., vol. 30, pt. II, Oct. 1951, pp. 1041–78.

2. Carrier Terminals for the L3 System, D. B. Penick. *Bell Laboratories Record*, New York, N. Y., vol. 33, no. 2, Feb. 1955, pp. 72–77.

3. L3 Coaxial System—Television Terminals, J. W. Rieke, R. S. Graham. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 541–51.

4. Switching at TV Operating Centres, A. L. Stillwell, A. D. Fowler. *Bell Laboratories Record*, vol. 34, no. 10, Oct. 1956, p. 366.

5. Wire Line Entrance Links for TD-2, J. W. Beyer. *Ibid.*, vol. 32, no. 3, Mar. 1954, pp. 108–10.

Error-Checking for 5-Channel Telegraphic Tape

R. A. BARBEAU
MEMBER AIEE

THE USES and the advantages of 5-channel paper tape are well recognized in the business world. Paper tape has been used for many years for message sending. Its use in telegraphic communication equipment serves as the life line of communications for many businesses. In recent years the decentralization of factory equipment and the general growth of businesses have made the transmission of accounting or scientific data between cities increasingly important. The rapidly expanding use of commercial computers and business machines will result in a further increase in the volume of data that must be communicated between cities. As more and more data are communicated, one becomes aware that a major improvement is required to refine the system, especially with regard to error handling. There is a practical limit on the improvements that may be made in transmission facilities to decrease the error rate. However, these improvements, plus a system of error detecting, could provide a highly efficient system that would be able to handle large volumes of data with a minimum amount of error-handling techniques.

This paper will describe an error detection system that is compatible with present-day telegraphic communication facilities, including automatic switching centers. It will describe a complete, practical system that has been tested and has proved to be highly efficient. The paper will discuss in detail the systems problems and engineering problems, and how they were approached.

Systems Problems

Systems used in the past have made use of group total control for the detection of errors. Information prepared at a point of origin, such as invoice writing, may be contained in a roll of tape 100 to 200 feet in length. At the point of origin a group total is compiled for the important items that should be checked, such as quantity of items or amounts of money. As shown in Fig. 1, this information is transmitted to a distant reception point, the conversion to other means of storage takes place, and it is followed by an attempt to balance to the control figures. Should the control totals fail to balance, there is at present no easy means of indicating or pinpointing where the error might have occurred in this group of records. It can, in fact, necessitate the reprocessing of a new tape at the point of origin for the full amount of information that did not balance. This is not only time-consuming and costly, but inefficient, because both correct and erroneous information is reprocessed.

It is known that there are five areas in which errors may occur, as follows:

1. The preparation of paper tape, especially when done by a manual method.
2. Reading of the paper tape for transmission.
3. Actual transmission of data over leased lines or other media.
4. Reperforation of paper tape at a distant point of reception.
5. Reading of tape for conversion.

To have a practical system of error detection, any rearrangement of codes must not have an adverse effect on communication equipment now in use, such as page printers or other subscriber sets. In fact, equipment already available in the field must be utilized. Rearrangement of codes to include checking must not make unintelligible or otherwise affect information printed on page printers. Basic telegraphic code arrangement rules must not be violated.

Any checking code must not interfere with switching centers now in existence. As an example, a figures-shift *H* character must not be used for record checking because this code has its primary use as a switching center control.

The present economy in transmission time must be maintained, which means that any added characters for checking must be kept to a minimum.

In a complete system that is to receive broad usage, any proposed device capable of producing checkable tape must accept electrical inputs from other business machines.

The problem of operator error with a manually keyed machine is probably one of the most difficult items to handle in the over-all system of error control. Many operator errors may be eliminated by a simplified error correction procedure and by a reduction in the number of operational decisions required of the tape punch operator.

Design Goals

After the systems problems mentioned had been reviewed, it was realized that design efforts should be concentrated in the areas of tape preparation and tape

---

Paper 58-2, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted July 12, 1957; made available for printing October 8, 1957.

R. A. Barbeau is with the International Business Machines Corporation, Poughkeepsie, N. Y.

The author wishes to express his appreciation to D. K. Close and P. M. Young for their helpful suggestions and criticisms in connection with the preparation of this paper.

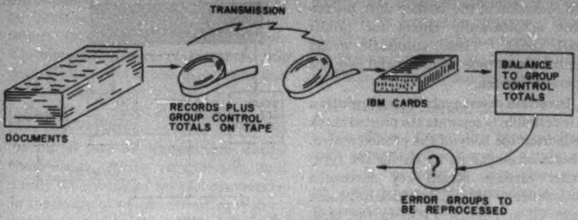

TRANSMISSION
DOCUMENTS
RECORDS PLUS GROUP CONTROL TOTALS ON TAPE
IBM CARDS
BALANCE TO GROUP CONTROL TOTALS
?
ERROR GROUPS TO BE REPROCESSED

Fig. 1 Data-flow chart

conversion. It also became evident that, to detect error at the converter, some type of "piggy-back" check should accompany the information in the form of a tape code.

The major goal was to design and supply an entire workable check system from start to finish, from the preparation of tape to the finished accounting medium, such as cards, for example.

Another important design goal was aimed at the operator preparing the punched tape. It was recognized that the operator plays a very important part in the accuracy of the over-all system. Consequently, any changes that would simplify the operator's work in the proper manipulation of keys, the reduction of key strokes, or other improved operating procedures would aid in producing a more accurate and better tape from an over-all systems standpoint.

Another objective was to design an error deletion routine that would be fast, automatic, and reliable, possibly controlled by a single keystroke. This would serve to relieve the operator of multiple decisions during the tape preparation.

The design of the system should keep total costs in line with the results gained. The system must contain the desired accuracy without imposing a hardship on the required transmission time, and it must not require changes in present telegraphic transmitting facilities.

The code arrangement, if changed to include checking, should not affect any of the existing systems in telegraphic communications. Particularly, it should not affect functions attributed to automatic switching centers. Also, such a code arrangement must be adopted that can make use of all telegraphic facilities.

Preparation of Checkable Tape

A variable length block of information was decided upon as a unit record of information. A unit record of this nature would fit a telegraph page printer, a typewriter, or a tabulating card. It was found that a unit record which corresponded to a line on a printed page was compatible for both error deletion and error detection.

ERROR DELETION

If deletion is to be automatic, an identification means must be used to signify the beginning of a record and again to indicate the point of error. The machine preparing the tape is basically a punching mechanism that does not contain a multiple-channel tape-reading station. It was decided to use a notch on the top edge of the tape for identification, which would allow for ease of sensing and would not require interpretation by a multiple-channel reading station. A notch is located between the last character of the preceding unit record and the first character of the new unit record. Another notch is used to identify the point of error.

If the operator recognizes a mistake during a given unit record, the depression of an error key notches the tape at the point of error. Automatic backing up of the tape to the point where the record started, as shown in Fig. 2, and automatic deletion or superimposing by the delete code will take place until a notch is read indicating the point of error. The notch in the tape will not affect the communication tape readers because its location is outside the operating area of the reading and run-out pins.

This manner of automatic error deletion fits a requirement of manual operation. No decisions are needed on the part of the operator other than to depress the error key at the time the error is recognized and to retype the line. The notch in the tape signifying a unit record is entered automatically at the beginning of every line. It is possible for the operator to make a correction for that line no matter where in the line she recognizes the error. The depression of the carriage return key may be likened to a motor bar depression at the end of an entry. The entry may be corrected at any time prior to the depression of the motor bar operation. This single decision by the operator has proved very effective and accurate in the preparation of paper tape.

CHECKING CODE

A decision was made to arrive at a piggy-back code by counting the code bits in the record. This code was then attached as part of the record to be transmitted. The decision to count the bits in a record blended with the fact that it could be directly coded to the figures-shift digits 1 through 9 and zero. It was decided that the check digit count would be the units position of the total count of the number of holes appearing in the tape following a beginning of unit record identification and including all of the end of record identification codes. It was not felt that a 2-column check code was required, and further tests have borne out the fact that the single-column check code is sufficient.

The generation of a check code by unit record signifies that a beginning of record identification must be used, as well as an end of record identification, because the same method of checking the information at the point of conversion must take place in the system. The identification of the unit record for the operation of a checking code cannot be the notch because the notch will never appear on the reperforated tape.

Because the letters shift code (LTRS)

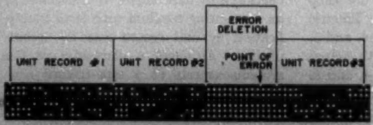


Fig. 2 (left). Error deletion

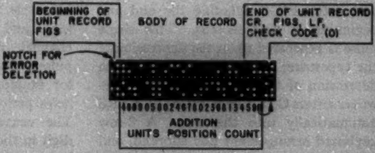

BEGINNING OF UNIT RECORD FIGS
BODY OF RECORD
END OF UNIT RECORD CR, FIGS, LF, CHECK CODE (0)
NOTCH FOR ERROR DELETION
000005002467802368134590
ADDITION UNITS POSITION COUNT

Fig. 3 (right). Unit record of tape

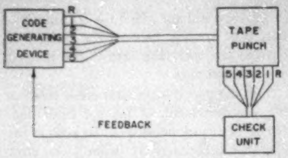


Fig. 4. Redundancy checking

is used for error deletion and as a run-in code, the figures (FIGS) shift code was selected to identify the beginning of unit record. The carriage return (CR) and line feed (LF) codes required for page printers were further expanded to provide an end of unit record identification. The check code requires a FIGS shift code, and the CR function requires a second nonprinting code. The identification of the end of unit record is then comprised of a 3-unit code group, namely, CR, FIGS, LF. The check code follows the end of unit record identification and becomes part of the unit record. A complete unit record of tape is shown in Fig. 3.

The FIGS entered between CR and LF becomes a very distinctive identification mark for the new checkable tape, as contrasted with a noncheckable tape containing CR, LF, in that sequence.

During the perforation of paper tape at the point of origin, the entry of the FIGS shift code at the beginning of a unit record conditions the machine to clear the adding counters, makes a reset zero check, and instructs the machine to add all information following this beginning of unit record identification. When the case of the next code is alphabetic, LTRS shift follows the first FIGS shift. While intelligence is being entered into the unit record of tape, addition of the holes making up the 5-channel code is in process throughout, including the end of unit record identification. The total units count of the number of holes will be recorded in a telegraphic code after the LF.

Automatic Cycles

Tape punch circuitry was designed to accept a single pulse in order to enter automatically CR, FIGS, LF, check code, and notch FIGS denoting the end and beginning of unit records. This single pulse is generated upon the depression of the typewriter carriage return key. The depression of the TAB key on the typewriter enters G, FIGS or FIGS, G, LTRS automatically into the tape. A 3-row keyboard arrangement was adopted, and the operation of the basket shift mechanism automatically enters the proper shift codes. These automatic cycles contribute to tape accuracy and ease the operator burden.

In certain cases involving page printers, it is desirable to separate the printed check code from the body of the printed record. This is accomplished by having the typewriter carriage return key depression automatically insert CR, FIGS, LF, check code, G, and notch FIGS in the tape. The G code operates the tabulating mechanism in the page printer. It is also desirable in some cases to insert LF codes following the beginning of record FIGS code. Subsequent depressions of the carriage return key insert the LF code only.

Redundancy Checking

Whenever two pieces of equipment are electrically connected, some sort of feedback is desirable to ensure the accuracy of both devices. It was decided that the code-generating device should furnish a redundancy 6-channel code, five of which were the telegraphic channels; see Fig. 4. Five telegraphic channels would be punched in the paper tape, and the sixth would operate a dummy punch. All six punches would operate circuitry to effect an odd check.

Circuits in the code-generating device are so designed that the failure of a single relay point or a single diode will cause no more than the omission of a single bit, or all bits, either of which omissions may be detected by the odd count redundancy type of check. This check takes place during the punching of every code in the tape.

Reliability Checks

In keeping with the accuracy of the system, reliability checks were added to ensure the earliest possible detection of machine malfunction and thus avoid lost time throughout the system.

As a supplement to redundancy checking, or when a code-generating device is not capable of supplying a sixth channel output, the design of the mechanism and circuitry provides a further accuracy check of the output medium and punching mechanism. On every tape-punching cycle a check is made between the input and the actual punching for each channel of tape information. Electric pulses, conditioning channels 1, 4, and 5, for example, are compared with actual holes perforated in channels 1, 4, and 5. The conversion of the decimal check digit in the counter to a telegraphic tape code is done on a 6-channel basis in keeping with redundancy checking.

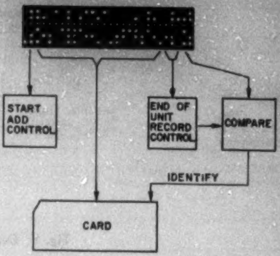


Fig. 5. Tape-to-card conversion

To assure the accuracy of the accumulated check digit, two sets of counters are used. These counters are fed by relays whose accuracy has been checked by previously mentioned methods. A comparison is made between these counters on each tape-punching cycle.

A second notch-reading pin is employed to check the accuracy of the first notch-reading pin. It provides assurance that a good record will not be overrun during error deletion. Both notch-reading pins are used to check tape movement at the beginning of every record.

When malfunctions are detected by these means, a signal locks the typewriter keyboard to prevent further processing of tape. The operator may then depress the error key, deleting the unit record, and attempt to retype the line. Thus unit failures are caught, and the operator is prevented from preparing faulty tape.

Paper Tape Transmission

The tape prepared as described may now be transmitted over commercial communication equipment, and reperforated tape prepared at the point of reception. It should be noted that this system does not require changes in standard equipment, and any improvements made in service will tend only to improve the system.

Error Detection During Conversion

Since the reperforated tape is checkable by unit record, its conversion to an accounting medium with error identification by unit record may be accomplished. Although a discussion of the systems aspects of a tape-to-card converter would supply a background for the following discussion, it is felt that the conventional converter functions might be indistinguishable from the detection means incorporated in the machine. Therefore this section of the paper will deal primarily with the error detection phase of conversion.

As previously listed in the systems problems section, errors can occur during the reading of tape for transmission, during transmission, during reperforation, and during the reading of tape for conversion. In the design of a tape-converting system, a careful review must be made of the types of errors that will be encountered, and enough intelligence must be built into the machine to insure fail-safe operation. This intelligence must also include reliability checks.

The error-detecting device attached to a tape-to-card converter contains primarily a single-position decimal counter under reset and add control of the first FIGS in a record. Accumulation of the code bits takes place in much the same way as in the device which prepared the checkable tape. The piggy-back check code in the tape is compared with the total in the counter under the control of the end of unit record codes, CR, FIGS, and LF. During tape reading, the analysis of the first FIGS code of a record resets the counter, makes a zero check, and conditions the counter to add succeeding code bits. The body of the record is punched into the card simultaneously with tape reading.

When the tape converter analyzes the end-of-unit record codes, CR, FIGS, LF, the machine is instructed that the next character to be read in the tape is the check code. During the reading cycle following the LF code, a comparison is made between the accumulated decimal count in the converter and the check code appearing in the tape. When the two numbers are in agreement, the converter gives a check signal for proper card identification. During the comparing cycle, if the digit appearing in the counter is in disagreement with the code which is read in the tape, the converter gives a noncheck signal for proper error identification of the card. Because the check code appearing at the end of a unit record is strictly for accuracy checking, it is not available for conversion purposes. When the figure shift G code has been inserted between unit records, the G is not made available for conversion purposes.

UNIT RECORD CONTROL

For efficient use of this type of check digit appearing at the end of a unit record, it is necessary to have close control over defining the record limits, both the beginning and the end of unit record.

The primary use of the beginning of record FIGS code is to control subsequent adding of the code bits of the unit record during conversion. To effect a control of a record that has a mutilated beginning of unit record FIGS code, the conversion of tape codes conditions an error memory device. The device supersedes check code comparison in order to effect error identification of this record.

Previous error-detecting experience led to setting up controls using all three end-of-unit record identification codes in a given sequence. When all three end-of-unit record codes are present, CR, FIGS, LF, in that order, the comparing cycle output depends upon agreement between the counters and the check code in the tape; see Fig. 5. If one of the three end-of-unit record codes has been mutilated, the error memory device is controlled and will cause the noncheck signal during the comparing cycle. Examples of these code mutilations follow:

Space, FIGS, LF
CR, V, LF
CR, FIGS, CR

The converter will acknowledge any two of the three end-of-unit record codes in correct sequence and cause comparing on the correct cycle. This comparing cycle control is not effective when an LF code follows the beginning of record FIGS.

If CR, LF appear to the converter in this order, they cause automatic comparing on the next cycle with an automatic noncheck signal output. This comparison serves two purposes:

1. Indication that noncheckable tape is being processed.
2. Prevention of switching directive codes from being processed as a correct record.

When the CR code is sensed not in its proper position, the error memory device will be operative; for example, if a channel 3 space code has been shifted to a channel 4 CR code in the body of the unit record.

Prior to testing this error detecting system, it was recognized that certain types of compensating errors (such as the dropping of one bit on one code and the adding of another bit on another code within a unit record) would not be detected solely by the check code appearing at the end of the unit record. It was also realized that certain types of transmission failures might result in the dropping or adding of ten bits of information. During the testing of this system, it was found that the compensating errors within a unit record were very infrequent, and the occurrence of an error of ten was extremely rare. The logic of the converter, prior to the application of error detection, could already take care of some of these compensating errors or errors of ten, and these checks were available at no extra expense to the system.

Conclusions

Numerous papers have been given on the subject of error detection, mentioning a check code. However, it is felt that for a system to be practical it should contain close limit control on the checking block. These controls make it possible to use a variable length record with a high efficiency of error detection. The system described here is aimed primarily at compatibility with the existing commercial communication facilities.

Conventional error handling in accounting systems is time-consuming and costly, because error detection is normally done by control groups at a point in accounting later than tape conversion. The checking system described in this paper allows for error detection by unit record during the conversion process. Reprocessing or retransmission of only the erroneous information saves considerable time. This system proves to be an accurate, practical checking means for paper tape.

Digital and Pictorial Photographic Electronic Recorder

R. G. McPHERSON
ASSOCIATE MEMBER AIEE

I. A. SONDERBY
NONMEMBER AIEE

A FILM DATA-RECORDING system is described which combines the advantages of pictorial and digital recording. This combination of photographic and digital techniques provides a record containing a picture of the event of interest with the adjacent frame used to record a digital code, which may be used to provide a record of instrumentation, time, or other identifying information.

The advantages of film recording include high density recording (systems of 6,000 bits per square inch are in operation with work reported on systems approaching 10^6 bits per square inch); high readout rates (systems of 200-kc bit rate are in operation and experimental 2-megacycle systems have been in operation); and high data storage per cubic foot, as demonstrated by microfilm files now being used. Automatic search through film reels based on recorded data is possible.

Data Recording

The upper limit of data-recording density on the film is determined by considering film tolerances; registration accuracy in recording and reading; effects of dirt, dust, and scratches; and complexity of readout and recording equipment. Experience with film reading projects indicates that data spots 10 mils by 10 mils are a practical size for the digital data. For 16-millimeter film this corresponds to rows of data of approximately 40 bits or 4,000 bits per inch of 16-millimeter film. With dimensions of 10 mils by 10 mils, successful data recorders and readers have been designed and operated. While it is certainly possible to provide a much greater density of data per square inch of film, the considerations of film dimensional stability with temperature, humidity, and age for systems intended for commercial type use indicate that these dimensions are practical limitations. In addition, for data rows in excess of 40 bits, the problems of providing a clock become somewhat more difficult.

Miniature incandescent lamps (0.069-inch diameter) have been used for the recording head light source. They are available from a number of companies and are a standard item in use for medical applications. Lamps of this type have been tested with over 4.6 million cycles of on-off operation without failure and continuous operation of over 800 hours without failure. The power requirement is 90 milliwatts at one volt. Requirements for this recorder system include analog-to-digital conversion with the digital output capable of providing one volt at 90 milliamperes. Examples of encoders that may be used include the commutating code wheel shaft position-to-digital code converter, of which a number of types exist. Relay devices may be used, as may transistor switches driven from a flip-flop register.

At The Magnavox Research Laboratories, several film data programs are in process or have been completed. One system recently completed provided for the recording of airborne instrumentation data consisting of up to 90 bits of data at rates up to 2,880 bits per second. In addition, pictorial data are collected to show the pilot's observations during the tests in question. The airborne equipment consisted of analog-to-digital conversion equipment, a timing information generator, and the controls necessary to start and stop recording based on events detected in the aircraft. The recording rates required were 16 and 32 frames per second.

Development work on design of a recording head included study of electroluminescent-cell light sources, small cathode-ray tubes, and miniature incandescent lamps. After consideration of the limitations of each of these recording methods and some experimental work, it was determined that the incandescent lamps recorder was the better choice for this application. Fig. 1 shows the recording head assembly used for this airborne application. The recording head contains 90 lamps and is approximately 1/2 cubic inch in volume. Fig. 2 shows the type of record obtained. For 16-millimeter film, one and one-half frames are used for picture area, one half a frame is used for data. The picture and digital data are recorded simultaneously through the shutter.

In the airborne recorder application three voltage signals, whose amplitudes were proportional to the quantity to be measured, were used to provide drive signals for three servo units. The commutating type of code wheels were used for the shaft position to digital conversion. This system provided a digital record on the film of ±15 degrees measured in 1-mil increments. In addition, measurements or indications were provided for the type of armament, time, slip or skid, prerun condition, and armament action. The airborne recorder is shown in Fig. 3.

Data Reader

Several methods of reading or extracting the information from the film were investigated. These methods are classified in three groups: optical-mechanical, cathode-ray tube, and storage tube. Each of these has definite advantages, depending on the application in question. Reading rates, tolerances, and reliability are important, and greatly affect the method to be considered.

Of the three types mentioned, the optical-mechanical method is perhaps the simplest in regard to method of readout. An experimental system was constructed using a rotating drum on which were mounted 15 mirrors, each consecutive mirror mounted at a slightly different angle from the previous one. The information to be read consisted of 15 rows of information with six bits in each row. The information itself was represented as 10-mil squares, separated by 10 mils. The field of information, 0.290 by 0.110 inch, was optically magnified ten times and projected on the face of the rotating mirrors. Each mirror in sequence would move its predetermined row of information across the face of a photomultiplier tube and appear at the output as a signal. The upper limits of read rates are determined by the speed of the rotating drum, the amount of information, the precision mounting of mirrors, the frequency limit of the sensing device, and mechanical tolerances.

For data-reading systems of greater density or for those systems requiring greater readout rates, the electronic type of readers have advantages over optical-mechanical systems. An example of this is the flying-spot scanner, a cathode-ray tube for providing a scanning raster

---

Paper 58-29, recommended by the AIEE Recording and Controlling Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 13, 1957; made available for printing November 1, 1957.

R. G. McPherson and I. A. Sonderby are with The Magnavox Research Laboratories, Los Angeles, Calif.

Fig. 1. Recording head assembly

with a multiplier phototube used to convert light changes to electric signal changes. The scanning raster may be made larger than the data field to be covered to allow for registration inaccuracies.

Storage tubes, such as the image-orthicon, the iconoscope, and the vidicon were studied for the film reader function. The iconoscope was made nearly obsolete with the development of the image-orthicon tube. This tube has a much higher sensitivity and efficiency rating compared to that of the iconoscope. In addition, the image-orthicon seemed unfeasible from the standpoint of price and because of its vulnerability to damage. The studies indicated that the vidicon storage tube would be the most practical of the three. Its cost is about $150, whereas the iconoscope ranged from $400 to $600 and the image orthicon was about $1,400. The vidicon is less susceptible to damage from scan failure or from overillumination; it is more compact and allows for more miniaturization in design.

A system using the vidicon as the film reading device is under development and preliminary tests are in progress at The Magnavox Research Laboratories. Tests indicate that a 40- by 60-bit array with each bit 10 mils square, is well within the limits of the reading device. The information matrix is projected on the photoconductive layer on the face of the vidicon. The photoconductive layer is then scanned from within, and signals corresponding to the light areas in the information array are presented at the output. These signals are amplified and are available for use in the decoding or sensing equipment. The vidicon is capable of a 600-line resolution. In this application a 290-line raster is used, which is within normal operation specifications. Effective spot size is approximately 2 mils, linearities in horizontal and vertical sweeps are 2%. Linearities of 1% may be obtained with a high-quality, specially designed yoke. The 40- by 60-bit array is scanned in 16 milliseconds. An overlap of 10% in both horizontal and vertical directions is used to relax tolerances caused by misalignment of the film. The useful area on the face of the vidicon is a rectangle 0.5 by 0.375 inch, requiring an optical reduction from film to vidicon. The resultant four scans per data row assure a preferred scan more nearly in the center portion of a row.

The reading rate for the system described is 150,000 bits per second. Higher rates could be achieved by reducing the size of the coded bits and increasing the number of bits in the array. The reader developed for the airborne recorder system automatically reads the film, frame by frame, and provides a printed record and punched-paper tape output. Readout was accomplished by using an electromagnetic short-persistance cathode-ray tube with a 100-line raster displayed on its face. The light from the flying spot which traces out this raster is focused on the film data area by an object lens which reduced the raster size by a factor of 10. The image of the raster exposed on the film was such that an overlap of the data area was provided in both the horizontal and vertical directions. This overlap takes care of registration inaccuracies which may be associated with the film transport mechanism. The light passing the data field in the scanning process is collected by a collector lens, and projected on the photocathode surface of a multiplier phototube. Thus, light of varying intensity is translated into a signal voltage of corresponding amplitude. The signal voltage so developed is amplified and fed into the system for decoding and printout. The reader is shown in Fig. 4.

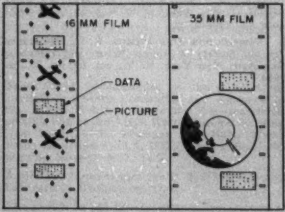


Fig. 2. Sample data record

Fig. 3. Airborne recorder

Fig. 4. Flying-spot scanner film reader

Scanning Patterns

For systems with short data rows a free-running clock of the pulsed oscillator or multivibrator type is adequate. By "clock" is meant the provisions of a gating waveform as interrogation pulse to sample the signals from the film data pickup device. This is necessary to identify adjacent bits on the film. Longer data rows (those of 10 bits or more) require either very tight tolerances on film dimensions and readout systems or a clock that is, in some way, synchronized with the film scanner.

Several methods of tieing in the clock with the film data are possible. A beam-splitting optical system may be used in which a fixed clock target is swept by the light source which simultaneously sweeps the data card. In this way the clock and data card signals are independent of time variations of sweeping circuitry. There is a space tolerance, however, that requires rather tight tolerances of the data film with respect to the fixed clock target. For the 10- by 10-mil case this registrator tolerance will approach one mil when the effects of film data dimensional change and recording location tolerances are included in the over-all problem. This method may be used only in systems which include a sweeping light source such as the cathode-ray tube reader. A self-clocking method using multivibrator circuits for the clock has been designed in which the film information changes are used to modify the clock rate or to shift the clock in a direction to resynchronize the clock. In this system, film information is compared with the clock pulse. If the clock is late, it is shifted to occur earlier. This method requires a digital code which ensures the occurrence of film data changes every few bits.

Several methods of scanning the film information have been studied; see Fig. 5. Of the various methods studied, the three which appeared feasible were those involving a full scan with a conventional raster such as used in the system designed and presently in use; a combination of partial scan and full scan; and a combination of vertical and horizontal scan. The full-scan method has been described. For the partial scan, two horizontal sweep rates would be employed, and the shorter sweep would continue until the timing mark was reached. The signal from the interception of the timing mark would enable the long or readout sweep to traverse the row of information and readout would occur. The shorter sweep would again take place and the same procedure would be repeated. An advantage of the partial scan method is that it would decrease the bit rate to approximately half that of a full-scan method, since there are approximately half as many total scanning lines per scanning raster. One disadvantage would be that two separate horizontal sweep systems would be necessary. They could be achieved with special sweep circuit and yoke design that would add to the complexity of a system.

The third method is to use a vertical scan which would continue to move in its vertical path until it intercepted a sensing tab. The signal derived from this interception would be used to enable the horizontal sweep, and readout of the information would occur. This method has one advantage over the previous two in that the bit rate would be approximately one third that of the full-scan method, and also less than that of the partial scan. There are added disadvantages, however, not only because special sweep-circuit and yoke designs are needed, but also because the trace would be required to return to within a close tolerance of its initial starting point. This tolerance would necessarily depend upon the application in question. Also, the long-term drift associated with tube and component decay, plus variations in operating voltage, would present a problem. However, these last two methods are feasible and have been considered from the standpoint of lowering the bit rates in systems with rates of one megacycle or greater, so that lower frequency circuits may be employed. Some means of identifying or sensing the the center portion of each row of information must be considered. A small square bit, half the size of the information bits, could be located directly in front of each row of information. If this small bit were intercepted, it would serve to enable the readout equipment to sense the information in that particular row. Another method would be to have the rows separated by a space and to utilize the techniques employed in the system previously mentioned. This would, however, reduce the amount of information to some extent, depending on the spacing required. A third method would be to employ a stripe with tabs projecting in the horizontal direction in line with each row of information present. The signal derived from the stripe, plus the tab, would serve to enable the readout circuit and cause the information to be sampled.

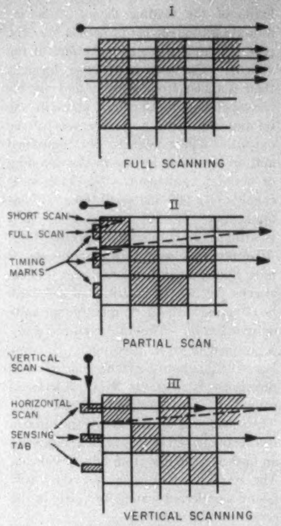


Fig. 5. Scanning patterns

Summary

The application of the type of equipment described may from the simple installation of code wheels attached to existing shafts on equipment, and camera with recording head, to the installation of nonloading, highly accurate (one part in 100,000) transducers with servo units and special timing generators. The film reader equipment may be designed simply to read the coded data and provide printed records, or it may include pictorial display by projection or by kinescope monitors remotely located. For data file systems, an automatic search feature may be provided, which on command, searches the film, frame by frame, until the desired information is located. Preliminary studies indicate that search of 40 frames per second of data, made up of 600 bits per frame, may be achieved with the use of techniques developed for existing machines. Even greater reading rates appear feasible with extensions of the existing methods.

The practical systems developed include a spot size of 10 by 10 mils, data rows of up to 40 bits, output rates of up to 200 kc, recording rates of 64 frames per second, with 90 bits per frame. Incandescent lamp recording has been used, and either cathode-ray tube or vidicon readers have been employed.

References

1. Television Engineering (book), D. G. Fink. McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

2. Standards on Television: Methods of Measurement of Aspect Ratio and Geometric Distortion. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, no. 7, July 1954, pp. 1098–1103.

3. Vidicon for Film Pickup, R. G. Neuhauser. *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 62, no. 2, Feb. 1954.

Bits of Information

A. S. ZAMANAKOS
NONMEMBER AIEE

Synopsis: The paper reviews the concepts of information, channel capacity, and equivocation. The probability of error is used to calculate the equivocation. It is then explained that by coding the original message or signal it is possible to incorporate into the coded message the error-detecting and error-correcting information. The requirements of an efficient code are stated, and the Shannon-Fano code cited as an example. The parity check is explained and the manner in which it makes use of the error-detecting and error-correcting information is illustrated.

THE FIRST STAGES of learning about a new field are usually fraught with difficulties of understanding and applying the various theorems and definitions. This paper will recount certain experiences in the process of becoming acquainted with communication theory, which is a subdivision of the broader field of information theory.

A measure of information was established by Hartley[1] and is defined by the equation

$$H' = \log N \tag{1}$$

where N is the number of signals, messages, symbols, or states available from a particular information source. If the base of the logarithm is equal to 2, then H' is in bits of information.

If the information source is an $m \times n$ matrix and each element of the matrix may assume L symbols, values, or levels, the measure of information is

$$H' = \log_2(L^{mn}) = mn \log_2 L \text{ bits} \tag{2}$$

A frame (or two fields) of a television picture is an example of such an information source; see Appendix I for a numerical example.

Equation 1 may be interpreted to mean that the frequency of occurrence of the various states which make up N is unknown; therefore they are all assumed to be equally probable. (It also happens that in this situation the information has been maximized.) If, on the other hand, something is known about the statistical nature of the information source, this must be taken into consideration. Assume that the simple probability (the appearance of one symbol or state, independent of the other symbols or states) $p(l_i)$, of each symbol L_i which the information source yields, is known. When these probabilities are applied to the matrix information source, and when mn is large, a particular symbol l_i will appear $mnp(l_i)$ times. Since the probabilities $p(l_i)$ are independent, each of the states represented by the various permutations of the mn symbols have the same probability of occurring. Therefore the probability of obtaining any state of the matrix (belonging to N) is

$$P_r = p(l_1)^{mnp(l_1)} p(l_2)^{mnp(l_2)} \dots p(l_L)^{mnp(l_L)} \tag{3}$$

and P_r is, for all purposes, a constant. Then $1/P_r$ is equal to the total number of states N in equation 1, see Appendix II. Substituting equation 3 into equation 1

$$\begin{aligned} H' &= \log_2(1/P_r) = -\log_2 P_r \\ &= -mn \sum_1 p(l_i) \log_2 p(l_i) \end{aligned} \tag{4}$$

and the information per matrix location or element is

$$H = \frac{H'}{mn} = -\sum_1^L p(l_i) \log_2 p(l_i) \tag{5}$$

Equation 5 was established by Shannon,[2] who called it the entropy of the information source.

If there is nothing to interfere with the communication of the output of the information source to some receiver,

$$H(x) = H(y) \tag{6}$$

where $H(x)$ is the entropy of the information source and $H(y)$ is the entropy of the receiver. This is a description of an ideal system. Because of thermal or resistor noise, it is certain that there will always be at least a minimum value of hindrance to any communication from source to receiver. Therefore, in place of the ideal situation described by equation 6, a definition is required of the joint entropy of the information source, the communication channel, and the receiver, as follows:

$$H(x,y) = H(x) + H_x(y) = H(y) + H_y(x) \tag{7}$$

The terms $H_x(y)$ and $H_y(x)$ are called conditional entropy or equivocation, and represent false information, i.e., $H_x(y)$ is a measure of the false information in the received information y, when the transmitted information is x; this is also true of $H_y(x)$ and y; see Appendix III.

When noise is apparent in the system, the simple probabilities introduced earlier are no longer applicable; instead, the joint and conditional probabilities, $p(x,y)$ and $p_x(y)$ or $p_y(x)$ must be employed. They are related as follows:

$$p(x,y) = p(x)p_x(y) = p(y)p_y(x) \tag{8}$$

As to the next step, it would appear reasonable to ask: "What is the rate at which we can transmit or communicate in the presence of noise?" According to Shannon the rate of transmission (of information) is equal to the entropy of the source minus the information lost due to noise:

$$R = H(x) - H_y(x) \tag{9}$$

In terms of reception

$$R = H(y) - H_x(y) \tag{10}$$

The maximum value R can assume is called the "channel capacity" and is designated by C:

$$C = \max[H(x) - H_y(x)] \tag{11}$$

In the noiseless case $H_y(x) = 0$ and C is evidently equal to $H(x)$ or $H(y)$, since $H(x) = H(y)$.

A value for C may be obtained in the noiseless case by using Shannon's original definition of channel capacity:

$$C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T} \tag{12}$$

where T is the maximum duration of the signals which comprise $N(T)$. For a teletypewriter system, under noiseless conditions

$$N(T) = (2^5)^{nT} \tag{13}$$

where n is rate of teletypewriter (character) transmission per unit time; therefore

$$C = \lim_{T \to \infty} \frac{\log_2 (2^5)^{nT}}{T} = 5n \tag{14}$$

If thermal noise which may be present is accounted for,

$$\begin{aligned} C &= \lim_{T \to \infty} \frac{\log_2 [\sqrt{(1 + p/\mathfrak{N})}]^{2WT}}{T} \\ &= W \log_2(1 + p/\mathfrak{N}) = W \log_2(p + \mathfrak{N}) - W \log_2 \mathfrak{N} \\ &= H(p + \mathfrak{N}) - H(\mathfrak{N}) \end{aligned} \tag{15}$$

Paper 58-34, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted April 2, 1957; made available for printing November 27, 1957.

A. S. Zamanakos is with the U. S. Naval Research Laboratory, Washington, D. C.

The author would like to express appreciation to S. F. George, consultant, Radar Division, Naval Research Laboratory, for his encouragement in this work.

Table I. Construction of Shannon-Fano Code

Example 1				Example 2			
Symbol No.	Probability	Code	Division	Symbol No.	Probability	Code	Division
1	1/4	0 0		1	1/4	0 0	
			(2)				(2)
2	1/8	0 1 0		2	1/4	0 1	
			(3)				(1)
3	1/8	0 1 1		3	1/8	1 0 0	
			(1)				(3)
4	1/8	1 0 0		4	1/8	1 0 1	
			(3)				(2)
5	1/8	1 0 1		5	1/16	1 1 0 0	
			(2)				(4)
6	1/16	1 1 0 0		6	1/16	1 1 0 1	
			(4)				(3)
7	1/16	1 1 0 1		7	1/16	1 1 1 0	
			(3)				(4)
8	1/16	1 1 1 0		8	1/32	1 1 1 1 0	
			(4)				(5)
9	1/16	1 1 1 1		9	1/32	1 1 1 1 1	

where W is the system bandwidth, in cycles per second, $P/\mathfrak{N}$ is the ratio of the average signal and noise powers, and the terms $H(p+\mathfrak{N})$ and $H(\mathfrak{N})$ correspond to $H(y)$ and $H_x(y)$ of equation 10. The term $\sqrt{(1+P/\mathfrak{N})}$ is the number of distinguishable voltage levels and $2WT$ is the number of elements or the length sequence of signals. The product $2WT$ is also the number of independent samples (or amplitudes) required to describe a function of duration T and bandwidth W.

Some of these relations will now be applied to a communication system composed of a 5 by 5 transmitting matrix and a 5 by 5 receiving matrix. The 25 elements or locations of the transmitting or information source are assumed equally probable of selection for (sequential) transmission. Therefore

$$H(x)=\log_2 25=4.64 \text{ bits per location} \qquad (16)$$

Assume in the transmission of a sequence of n signals that there will be a total of e errors, and that the location of these errors is known. This means that, to correct each error, only the remaining 24 possible elements or symbols need be considered. This selection represents an entropy of $\log_2 24$. Since the sequence length is n and there are e errors, a total of $e \log_2 24$ bits of information, or an average of $e/n \log_2 24$ bits per symbol, are needed to correct the errors. From the definition of equivocation,

$$E_1=p_e \log_2 24 \qquad (17)$$

which is the equivocation per received symbol and where

$$p_e=\frac{e}{n} \qquad (18)$$

is the probability of error.

But if the location of the errors is unknown, additional information is needed in order to establish their location. Using p_e as the probability of error and q as the probability that the symbol is correct yields, from equation 3

$$P_r=p_e{}^{np_e}q^{nq} \qquad (19)$$

The average amount of information per symbol required to locate the errors is

$$E_2=-\frac{1}{n}\log_2 P_r$$

$$=-(p_e \log_2 p_e+q \log_2 q) \qquad (20)$$

Since entropy has the additive property, the total equivocation of the system per symbol is

$$E=E_1+E_2=-(p_e \log_2 p_e+q \log_2 q)+p_e \log_2 24 \qquad (21)$$

Equation 21 represents the amount of additional information per symbol required to locate and correct the received message. E may be calculated if certain signal shape and thermal noise are assumed; see Appendix IV for another example. The signal shape may be a prescribed pulse (or no-pulse) occurring at specified times.

The way to calculate the measure of information and equivocation of a noisy system has been considered. The next step is to consider the manner in which this error-detecting and error-correcting information may be conveyed or communicated. This is to be accomplished by the use of error-detecting and error-correcting codes. Such a technique is the "parity check" and is quite effective in systems employing binary coding: 0's and 1's, or pulse and no-pulse. The parity check method will be touched upon. A rigorous presentation may be found in the work of Hamming[3] and of Gilbert.[4] More complex codes have been proposed by Silverman and Balser,[5] and others.

But before proceeding let us just mention the efficiency aspects of coding. A measure of efficient coding[6] is one which yields for the different symbols l_i and corresponding durations t_i the most probable a priori density distribution, namely the following:

$$\frac{-\log_2 p(l_i)}{t_i}=\beta \text{ (a constant)} \qquad (22)$$

where $p(l_i)$ is the proportion of times that the symbol l_i appears in the message. Equation 22 is intuitively satisfying for it may be interpreted as follows: The time required to communicate a particular symbol shall be inversely proportional to its frequency of occurrence or appearance in the message. This criterion was closely approximated in establishing the Morse code since the frequently used letters were assigned the shortest symbols.

The Shannon-Fano code is an example of an efficient code.[7] In it the length of the code group, in binary digits, is set approximately equal to $-\log_2 p(l_i)$ where $p(l_i)$ is the probability of the symbol l_i. The code is simple to construct and is as follows: The symbols available at the information source are set down in a column of decreasing probability. The column is divided into two divisions, upper and lower, in such a manner that the sums of the probabilities of each division are approximately equal. Since it will not always be possible to make the sums of each of the two resulting divisions equal, it will be agreed in advance that the probability sum of the upper division will be greater than that of the lower. The sums resulting from the first division process will be equal to 1/2. The digit 0 is assigned as the first digit in the code group for all the symbols appearing in the upper half and the digit 1 is assigned as the first digit in the code group for all the symbols appearing in the lower half or division. The division process is repeated, but now the probability sums are equal to 1/4. Again, the digit 0 is assigned, but as the second digit in the code group, for all the symbols which appear in the upper half and the digit 1 for all those in the lower. This division process is repeated until there is a single member only in each division.

An essential property of the Shannon-Fano code is that no short code group can appear as the initial portion of a longer code group; thus a built-in code group recognition aid is provided. Table I is an illustration of this process.

To return to the subject of parity checks, the established rules (which appear to be standard) are: 1. A sequence of checking signals will be sent, along with the information signals, and together they will constitute a code group; the location of the checking signals is fixed; 2. Adding the information signals and the checking signals in a prescribed manner will yield an even-valued number; since binary digits are used, the sum must terminate in a zero, in other words, an even number of 1's.

An easy example to analyze would be to send one information signal A and two checking signals B and C, and to specify that

$$A+B=\text{even} \tag{23}$$

$$A+C=\text{even} \tag{24}$$

A would be correct under these conditions. Assume that there is only one error; then, if equations 23 and 24 are both odd, it is known that A as received is incorrect. If one equation is even and the second odd, then the checking signal associated with the odd sum is incorrect.

The appearance of two actual errors would lead to wrong conclusions. Fortunately, including one additional checking signal D would determine if two errors were present. The additional checking equation would be

$$A+D=\text{even} \tag{25}$$

For example, if $A=1$ and $B=1$, when all the information and checking signals should be equal to zero, the following would prevail:

$$1+1=\text{even } (=10) \tag{23A}$$

$$1+0=\text{odd } (=1) \tag{24A}$$

$$1+0=\text{odd } (=1) \tag{25A}$$

and it would be known that there were two errors, but their location would not be known.

Returning to the situation in which no more than a single error can occur in a sequence of message symbols, the error may be located if enough checking signals are employed. Using a large enough signal-to-noise ratio[8] would ensure that only a single error would be likely. If there are m information signals and k checking signals, $(m+k+1)$ situations may be described as: 1. all signals correct, or 2. one of the m information signals or one of the k checking signals incorrect. In general, if the total number of transmitted symbols (or signals), composed of information and checking signals, is equal to $m+k$, it is seen that there are L^{m+k} possible transmission states. Out of this number of states L^k represents the various states which the sequence of checking signals may assume. Because binary digits are used, and $L=2$, the relation

$$2^k \geq m+k+1 \tag{26}$$

must hold. The equality relation represents the most efficient checking procedure. But even with the equality condition it may be that unavoidably, more bits of checking information than necessary are being sent. For example, assume the following:

1. A zero has been received.
2. The probability that a zero was sent is $q_0=0.995$.
3. The probability that a 1 was sent is $pe_0=0.005$.

If a 1 was received, the probability that a 1 was sent is $q_1=0.995$ and that a zero was sent is $Pe_1=0.005$. The equivocation per symbol (using equation 37) is $H_y(x)=0.045$ bits per symbol. In this example, two bits of checking information are used; therefore 2/0.045, or almost 45 times more checking information, has been included than theoretically necessary.

If there are two information signals, m_1 and m_2, it is known from equation 26 that three checking signals, k_1, k_2 and k_3 must be employed. The checking sums, established before hand, would be

$$m_1+k_1=\text{even}$$
$$m_2+k_2=\text{even}$$
$$k_1+k_2+k_3=\text{even}$$

and it would be determined as to which of the five signals was incorrect or if they were all correct. The number of checking sums or equations is equal to the number of checking digits, see Appendix V.

Of course, this error-detecting and error-correcting process implies a certain disadvantage, usually an increase in system bandwidth when the coding characters are transmitted at a constant rate. For example, in a teletype system if a bandwidth W is required to transmit the five information signals, it is calculated, from equation 26, that at least four checking signals are required. Thus the code group is composed of nine signals. If it requires time

$$t_0=\frac{5}{2W}=\text{a constant} \tag{27}$$

to send the five information signals, then, in order to send the nine signals in the same length of time, i.e.,

$$t_0=\frac{9}{2}W' \tag{28}$$

the bandwidth required is

$$W'=\frac{9W}{5} \tag{29}$$

In other words, the bandwidth has been increased by a factor equal to the ratio of the number of signals in the code group to the number of information signals. This shortening of the signals renders the system more susceptible to errors (because of the increased bandwidth), and an increase in the signal-to-noise ratio may be required to offset this condition. An additional task will be to decide if this is too great a price to pay for relatively error-free communication.

Appendix I. Information Content of a Television Picture, Assuming Zero Redundancy

If the horizontal scan frequency is 15,750 cycles per second, the period is equal to 63.5 μsec (microseconds). The sync and blanking consume 10.3 μsec, leaving 53.2 μsec as the visible scanning time t_v, in a line or row. With a 4-megacycle bandwidth W, the shortest picture element in a line which may be distinguished is $1/2W$ seconds' duration. The number of visible elements in a line will then be $t_v/(1/2W)=425.6$. Ordinarily 40 lines are lost because of the time required for retracing and blanking between fields; therefore, only 485 out of 525 lines are active. Assume that there are eight levels of intensity or picture element brightness. Substituting the numbers into equation 1 yields

$$\begin{aligned} H' &= \log_2 (l^{mn}) \\ &= \log_2 (8^{525\times 485}) \\ &= 76{,}875 \text{ bits per picture} \end{aligned} \tag{30}$$

Appendix II

The question has been raised as to whether N (or $1/P_r$) actually approximates the total number N_T of matrices having mn elements. The following is offered as proof that it does.

Assume that the total number of matrices containing mn elements is N_T. Out of this number of matrices there will be N of them containing $mnp(l_1)$ symbols of type l_1, and $mnp(l_2)$ symbols of type l_2, and so on to symbol of type l_L. The probability a matrix belonging to the set N is $P_r'=N/N_T$ and the probability of obtaining a particular matrix from N_T but belonging to the set N is

$$\begin{aligned} P_r &= p(l_1)^{mnp(l_1)}\, p(l_2)^{mnp(l_2)} \ldots p(l_L)^{mnp(l_L)} \\ &= \prod_1^L p(l_i)^{mnp(l_i)} \end{aligned}$$

With the description of the matrices belonging to the N, the latter may be found, i.e.,

$$\begin{aligned} N &= \frac{(mn)!}{[mnp(l_1)]! \ldots [mnp(l_L)]!} \\ &= \frac{(mn)!}{\prod_1^L [mnp(l_i)]!} \end{aligned}$$

Now, for P_r', as a result of this,

$$P_r'=NP_r=\frac{(mn)!}{\prod_1^L [mnp(l_i)]!} \prod_1^L p(l_i)^{mnp(l_i)}$$

Taking the logarithm and using Stirling's approximation yields

$$\log P_r' = \log (mn)! - \log \prod_1^L [mnp(l_i)]! + \log \prod_1^L p(l_i)^{mnp(l_i)} \cong$$

$$(mn)\log(mn)-\Sigma[mnp(l_i)]\times$$
$$\log[mnp(l_i)]+\Sigma[mnp(l_i)]\times$$
$$\log p(l_i)\cong(mn)\log(mn)-$$
$$(mn)\log(mn)\Sigma p(l_i)-(mn)\Sigma p(l_i)\times$$
$$\log p(l_i)+(mn)\Sigma p(l_i)\log p(l_i)\cong 0$$

Therefore $P_r'=NP_r\cong 1$ or $N\cong 1/P_r$ and, because $P_r'=N/N_T$, it follows that $P_r=1/N_T$. Thus $N\cong N_T$, which was to be proved. The matrices dealt with here are of the type that Shannon specified in his theorem 3.[3]

Appendix III. Equivocation of $H_x(y)$

If $H(x,y)$ represents the entropy of a joint event, the transmission of x and the reception of y, and from it is subtracted the entropy at the transmitter, there remains the undesired entropy due to noise or interference. This undesired entropy is, of course, false information, and will be called equivocation. To equate,

$$\begin{aligned}H(x,y)-H(x)&=-\sum_{i,j}p(i,j)\log p(i,j)+\\&\qquad\sum p(i,j)\log p(i)\\&=-\left[\sum_{i,j}p(i,j)\log\frac{p(i,j)}{p(i)}\right]\\&=-\left[\sum_{i,j}p(i,j)\log p_i(j)\right]\\&=H_x(y)\end{aligned}\tag{31}$$

The following relations have been used:

$$\sum_j p(i,j)=p(i)\tag{32}$$

$$p(i,j)=p(i)p_i(j)\tag{33}$$

Appendix IV. Probability of Error[8] and Equivocation

A communication system (bandwidth W) is sampled at the prescribed intervals. The system employs pulses whose maximum amplitude at the prescribed sampling points is equal to v_0 if there is no noise present. But since thermal noise will be present a signal threshold must be established, which will be equal to $v_0/2$; i.e., if the sample value exceeds $v_0/2$ it may be said that a pulse is present; if the sample value is less than $v_0/2$ it may be said that no pulse is present. Since only the presence of thermal noise and positive-valued pulses at the sampling points is assumed, the probability that the noise will exceed $v_0/2$ is

$$p_e=\frac{1}{\sigma\sqrt{2\pi}}\int_{v_0/2}^{\infty}\epsilon^{-\frac{t^2}{2\sigma^2}}\,dt\tag{34}$$

where the variance σ^2 is the mean noise power. Making a change of variable, $\lambda=t/\sigma$ yields

$$p_e=\frac{1}{\sqrt{2\pi}}\int_{v_0/2\sigma}^{\infty}\epsilon^{-\frac{\lambda^2}{2}}\,d\lambda\tag{35}$$

As a result of this change of variable the lower limit of integration is a signal-to-noise ratio which may be put in the form

$$\frac{v_0^2}{\sigma^2}=\frac{p_s}{\mathfrak{N}}\tag{36}$$

where p_s is the signal pulse power and $\mathfrak{N}$ is the noise power.

For example, if the signal-to-noise power ratio, $p_s/\mathfrak{N}$ is 22 decibels, then $v_0/2\sigma=6.294$. In the Work Projects Administration Tables of Probability Integrals,[9] a value is given of $1.59\times1.59\times10^{-10}$ for the probability of error p_e. The complement of p_e is $q=1-p_e$, which is the probability that the symbol is correct. Using the general equation of total equivocation

$$E=-(p_e\log_2 p_e+q\log_2 q)+p_e\log_2(N-1)\tag{37}$$

determines the additional information (check signals) required to correct the errors. N in equation 37 has the same significance as in equation 1. Equation 37 also reveals one of the advantages of using a binary signal (or information) transmission system. This advantage is that $p_e\log(N-1)$ is equal to zero since $N=2$. When the total equivocation is decreased, it is therefore unnecessary to send as much additional information.

Appendix V. General Process

The process used to determine the checking sums was one of deduction, i.e., first by using k checking sums the correctness of the first information signal is established; then by taking $(k-1)$ checking sums, the correctness of a second information signal can be determined. This

k

step is performed $(k-1)$ times where

$$\binom{k}{r}=\frac{k(k-1)\ldots(k-r+1)}{r!},\ \binom{k}{0}=\binom{k}{k}=1$$

and the range of the index is $0\leq r\leq k$. Then $(k-2)$ sums, etc., are taken. In the foregoing, similarity is noticeable to the problem of algebra concerning the number of ways in which k elements (in this instance, checking sums) may be grouped, i.e., k at a time, $(k-1)$ at a time, $(k-2)$ at a time, etc. Thus

$$\binom{k}{0}+\binom{k}{1}+\ldots+\binom{k}{r}+\ldots+\binom{k}{k}=N\tag{38}$$

as the total number of different groups (combinations). Next, it may be recalled from the binomial theorem that

$$(a+b)^k=\binom{k}{0}a^k+\binom{k}{1}a^{k-1}b+\ldots+\binom{k}{r}a^{k-r}b^r+\ldots+\binom{k}{k}b^k\tag{39}$$

and writing $a=b=1$ yields

$$2^k=\binom{k}{0}+\binom{k}{1}+\ldots+\binom{k}{r}+\ldots+\binom{k}{k}=N\tag{40}$$

or

$$2^k=1+k+\ldots+\binom{k}{r}+\ldots+1\tag{41}$$

Using the equality sign in equation 26 and solving from m yields

$$m=2^k-k-1\tag{42}$$

Combining equations 41 and 42 yields

$$m=2^k-k-1=1+k+\ldots+\binom{k}{r}+\ldots+\binom{k}{k-2}\tag{43}$$

Recalling the order in which the correctness of the information signals was established, equation 43 may be interpreted to mean that the k checking sums establish the correctness of the first information signal, and that

$\binom{k}{k-1}$ checking sums

establish the correctness of the next k information signals as indicated by the second term on the right-hand side of equation 43; or, in general, the $(r+1)$ term on the right-hand side of equation 43 determines the correctness of the next

$\binom{k}{r}$ information signals.

Appendix VI. Hamming Codes

Single-Error Detection

A single error may be detected by adding a (checking) binary digit to the message sequence of binary digits so that there are an even number of 1's in the resulting code group. This is called an even parity check. The combination of the information and checking digits is called a code group.

Single-Error Correction

The number of k checking digits required is determined by the relation $2^k\geq m+k+1$, where m is the number of message digits. Hamming determined the checking digit locations in the code group to be 1, 2, 4, 8, ..., numbered from the left of the code group. The first parity sum checks locations 1, 3, 5, 7, ...; the second checks locations 2, 3, 6, 7, ...; the third checks locations 4, 5, 6, 7, ...; Notice that the binary representations of the locations checked by the ith parity sum have a 1 in the ith location. The numbering in this instance starts at the right of the binary digit representation.

If the first parity check fails, a 1 is recorded (or a zero if it is satisfied). If the second parity sum fails, a 1 is again recorded but placed to the left of the binary digit recorded from the previous check. This process is performed k times and therefore yields a k-digit "checking number." The decimal representation of the checking number is the location of the (single) error in the code group. For example, if the number is zero it means that no error is present. If the number is 3 (011 in binary form) it means that the binary digit in location 3 of the code group is in error, etc.

Single-Error Correcting and Double-Error Detection

This process is accomplished by adding one more checking digit to the code group used in simple-error correction, and corresponds to location number $(m+k+1)$. The additional parity check which is introduced requires that the new code group (composed of $(m+k+1)$ binary digits) be even.

The following cases may exist:

1. No error. All the parity checks are satisfied.

2. Single error. The last parity check (which is applied to the new code group) fails, and the checking number obtained by single-error correction locates the error. The checking number 0 means that the last location is in error.

3. Two errors. The last parity check is satisfied but the checking number obtained by single-error correction is not equal to zero.

References

1. TRANSMISSION OF INFORMATION, R. V. L. Hartley. *Bell System Technical Journal*, New York, N. Y., vol. 7, July 1928, pp. 535–64.

2. MATHEMATICAL THEORY OF COMMUNICATION (book), C. E. Shannon, W. Weaver. University of Illinois Press, Urbana, Ill., 1949.

3. ERROR DETECTING AND ERROR CORRECTING CODES, R. W. Hamming. *Bell System Technical Journal*, New York, N. Y., vol. 29, no. 2, Apr. 1950 pp. 147–60.

4. A COMPARISON OF SIGNALLING ALPHABETS, E. N. Gilbert. *Ibid.*, vol. 31, no. 1, May 1952, pp. 504–22.

5. CODING FOR CONSTANT-DATA-RATE SYSTEMS, PART I, R. A. Silverman, M. Balser. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Sept. 1954, p. 1428; PART II, vol. 43, June 1955, p. 728.

6. SCIENCE AND INFORMATION THEORY, L. Brillouin. Academic Press, New York, N. Y., 1956, p. 71.

7. EFFICIENT CODING, B. M. Oliver. *Bell System Technical Journal*, vol. 31, no. 4, July 1952, pp. 724–50.

8. THE PHILOSOPHY OF PCM, B. M. Oliver, J. B. Pierce, C. E. Shannon. *Proceedings*, Institute of Radio Engineers, vol. 36, Nov. 1948, pp. 1324–31.

9. TABLES OF PROBABILITY INTEGRALS. Work Projects Administration, Washington, D. C., vol. II, 1942.

Study of Return Losses on Loaded Trunk Cables and Methods Taken to Improve Them

G. H. SPEAKE
ASSOCIATE MEMBER AIEE

RETURN loss as a measure of a circuit's uniformness has been used in the Telephone Industry for many years. It usually determines the amount of gain which may be introduced into a circuit before encountering undesirable conditions of echo or singing. While the toll plant, or long distance circuits, have been of primary interest, today's transmission objectives tend to emphasize the need for accurate analysis of all trunk cables with a view toward their improvement. This paper will review briefly the meaning of return loss, and attempt to acquaint the reader with some of the practical factors encountered in analyzing and improving the return loss of loaded trunk cables.

Return Loss

IMPEDANCE IRREGULARITIES

The return loss at an irregularity in a loaded cable has been defined in Bell System Practices and other literature as the transmission loss corresponding to the absolute ratio of the current, that would flow if no irregularity were present, to the amount of current reflected by the irregularity. A determination of its value at the point of irregularity involves a consideration of three impedances. If Z_{ox} is the iterative impedance to the left of a point, $Z_o\,(1-x)$ the iterative impedance to the right of the point, and Z_t the impedance looking through the irregularity, the return loss in db (decibels) is given by

$$RL = 20\log_{10}\frac{Z_{ox}+Z_t}{Z_{ox}-Z_t} + 20\log_{10}\frac{Z_0(1-x)+Z_t}{Z_{ox}+Z_t} \quad (1)$$

For irregularities located at mid-coil or mid-section, the second term reduces to zero, since the iterative impedance to the left and right are equal. For all practical work, the second term is usually ignored with negligible error. The return loss as determined may be referred to some other point in the circuit by adding twice the value of the intervening transmission loss.

STRUCTURAL IRREGULARITIES

While it is impractical to calculate the expected structural return loss of an individual cable pair before installation, it is feasible to determine a value of return loss which should be equaled or exceeded by any stated percentage of pair within a cable sheath, or a group for which the controlling parameters are known. The procedure derived by Crisson in 1925 is in use today.[1] For any loaded cable

$$S_1 = S_H + S_W + S_F - S_A + (S_n - S) \quad (2)$$

S_1 is the return loss in db which, it is expected, the selected percentage of pairs will equal or exceed, S_H is known as the irregularity function, S_W the frequency function, S_F the distribution function, S_A the attenuation function, and $(S_n - S)$ is a correction applicable to circuits which are not long enough to act as infinite lines.

In equation 2, S_H is primarily the factor which affects S_1. It is equal to 20 $\log_{10} 1/H$ where H is the square root of the sum of the inductance deviation squared plus the capacitance deviation squared. It is expected that the reader is either familiar with this concept or that the reference will be consulted for a thorough discussion.

CAUSE OF IRREGULARITIES

Equations 1 and 2 are presented here to show that the return loss of the loaded trunk is dependent upon the structural uniformness of the cable, and any additional impedance irregularities which may exist. From equation 2, it is seen that structural irregularities are dependent on the variations in manufacture, placing, and splicing of the cable, and also on the variations in the manufacture and spacing of the loading coils. Impedance mismatches which occur in trunks composed of different-gauge cable, while not considered in equation 2, are irregularities,

---

Paper 58-214, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 15, 1957; made available for printing December 17, 1957.

G. H. SPEAKE is with the Southern Bell Telephone and Telegraph Company, Birmingham, Ala.

The author wishes to express his appreciation to plant and engineering associates for assistance in securing data for this paper. Credit is especially due to J. W. Joyner and I. N. Howell, Jr., for their aid in making measurements and compiling results.

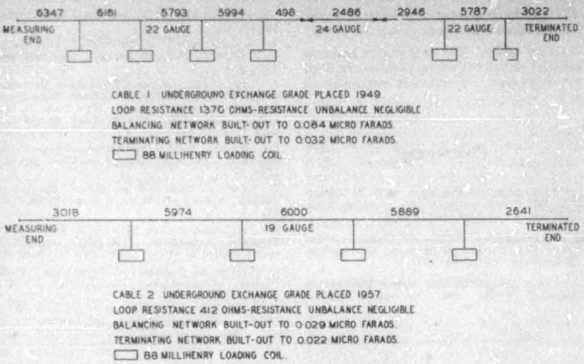

6347
6161
5793
5994
498
2486
2946
5787
3022
MEASURING END
22 GAUGE
24 GAUGE
22 GAUGE
TERMINATED END
CABLE 1 UNDERGROUND EXCHANGE GRADE PLACED 1949
LOOP RESISTANCE 1370 OHMS-RESISTANCE UNBALANCE NEGLIGIBLE
BALANCING NETWORK BUILT-OUT TO 0.084 MICRO FARADS.
TERMINATING NETWORK BUILT-OUT TO 0.032 MICRO FARADS.
88 MILLIHENRY LOADING COIL.
3018
5974
6000
5889
2641
MEASURING END
19 GAUGE
TERMINATED END
CABLE 2 UNDERGROUND EXCHANGE GRADE PLACED 1957
LOOP RESISTANCE 412 OHMS-RESISTANCE UNBALANCE NEGLIGIBLE
BALANCING NETWORK BUILT-OUT TO 0.029 MICRO FARADS.
TERMINATING NETWORK BUILT-OUT TO 0.022 MICRO FARADS.
88 MILLIHENRY LOADING COIL.

Fig. 1. Layout data for Cable 1 and Cable 2

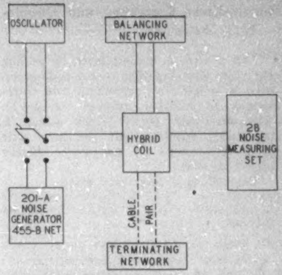

OSCILLATOR
BALANCING NETWORK
HYBRID COIL
28 NOISE MEASURING SET
201-A NOISE GENERATOR 455-B NET
CABLE PAIR
TERMINATING NETWORK

Fig. 2. Arrangement of test equipment for measuring echo return loss or return loss at single frequencies

and could be expected to affect the return loss. From the practical standpoint, the operating engineer has the most control over placing and splicing of particular cables and the spacing of loading coils. Of course, he also may choose proper terminations to reduce impedance mismatches at terminal points.

Existing Plant

TERMINATIONS

Although this paper is primarily concerned with trunk cable, the termination necessary to complete a circuit must of course be considered, since equations 1 and 2 show that the total return loss is dependent upon the terminal impedance match and the structural uniformness of the cable. The signalling, idle circuit, and talking or through conditions must be considered separately, because the proper operation of the circuit may require different terminations in each of these cases. For the talking or through conditions, the circuit termination normally consists of a repeating coil or transformer, the exchange loop cable, and the subscriber instrument. The loop cable plant generally consists of the most economical combinations of cable gauges, chosen within the limits of resistance design,[2] which combinations naturally present different impedances for different loops. Therefore, the choice of a transformer to match the trunk to the loop usually must be based on some average impedance value. The impedance of the telephone set is also fixed by design at an average value. For those trunks which are equipped with repeaters or gain devices, idle circuit terminations or a repeater disabler may be provided for the idle circuit or signalling conditions.

LOADED TRUNKS

The type of cable normally used for trunk circuits consists of paired conductors having various lengths of twist and a mutual capacitance of approximately 0.082 microfarad per mile. This is generally referred to as "exchange grade" in comparison with a toll grade or quadded conductor cable. The type and spacing of loading coils are recorded, but available records usually give little or no information about details of splicing or manufacture irregularities and tolerances. Therefore, for calculations, it is necessary to assume values for the unknown factors based on average figures and a knowledge of the splicing procedures which have usually been employed. "Bell System Practices" gives 0.007 as the standard deviation for 88-

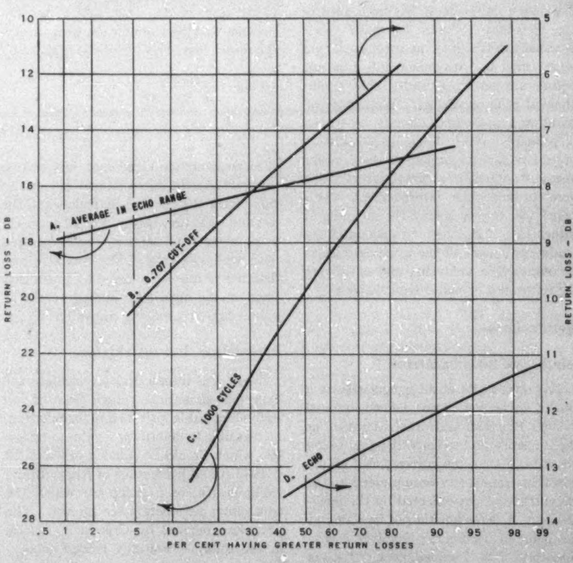


Fig. 3. Distribution of return losses for Cable 1

millihenry coils, and 0.027 as the loading-section capacitance deviation for exchange grade cable loaded at 6,000-foot intervals without deviation test splicing or reel allocation. The inductance deviation due to loading-coil spacing may vary from 0.005 to 0.03 or more. Thus the return loss may be greatly influenced by coil spacing irregularities.

A survey has shown that, in one metropolitan area having approximately 38,000 trunk pairs, 22% of the total are composed of cable pairs of two or more gauges. Trunk pairs in this case are understood to be those pairs between main frame terminations. Since these pairs are sometimes strapped together at several main frames in order to make up a complete trunk, at least 22% of the total trunks must have a split-gauge makeup. Also, approximately 48% of the trunk pairs were in the same cable sheath as pairs used for exchange loops. Although this situation does not affect the return loss of the trunks, it does make it difficult to determine the return loss distribution of the cable by measurements, because exchange loops are not readily available for testing.

Improvement Considerations

General

Because earlier transmission objectives for most trunks were from 8 to 10 db, the return loss was usually not very critical. However, with today's objectives of via net loss to 6 db, and even in some instances of zero loss or a net gain, the trunk cable must in some cases meet the same stringent requirements as an intertoll trunk. Preliminary surveys show that many of the existing trunk cable pairs do not have high enough return loss to support the gain necessary even for a 6-db trunk. Where the percentage of existing pairs that are capable of supporting the required gain is insufficient, it will eventually be necessary either to place new cable, spliced for a high return loss, or to improve the return loss of the existing pairs.

Although it is usually rather difficult and expensive to apply available methods of improvement to working cables, there are several methods which, if properly applied, should result in a satisfactory return loss for cable of the type normally used. The expected return loss for various conditions may be calculated, and the most economical choice of improvement methods which will produce the desired results may be determined. Costs may vary considerably, depending on the particular situation.

As some areas have not required acceptance tests on loaded trunk cables in the past, it may be advisable before attempting any improvement of an existing cable first to determine whether some trouble condition, such as a missing loading coil or bridged tap, is contributing to a poor return loss. Any such condition should of course be cleared before proceeding with improvement plans.

For improvement, the loading coils may be more accurately spaced, the cable may be randomed, or capacitance deviation test spliced in various lengths, solid gauge pairs may be provided, and all splices soldered. Also, building out of loading sections and end sections with stub cable or commercial-type capacitors may be necessary. In addition, impedance compensators may be used at terminal points.

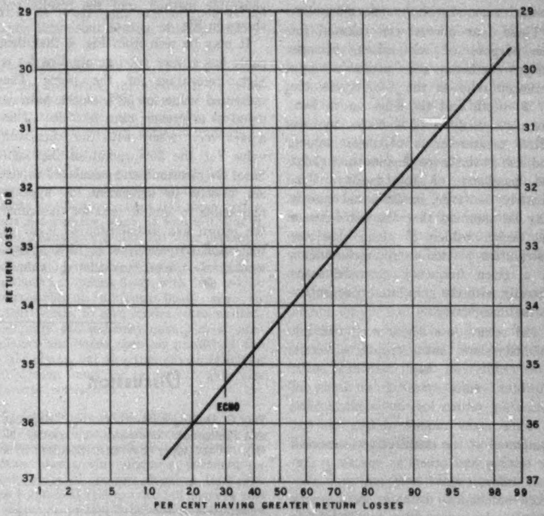


Fig. 4. Distribution of return losses for Cable 2

Other methods may also suggest themselves in particular cases. For example, it was found necessary in one instance to rework a small, 51-pair 22-gauge cable about 10 miles in length which had been hurriedly spliced in order to meet a service date. The cable was reterminated in a jack field, and patching cords were provided to maintain service during the splicing operations. The capacitance of each pair in each loading section was measured, and then all pairs were built out with commercial-type capacitors to a value equal to the highest capacitance found. This sufficiently improved the structural uniformness so that the repeaters could be satisfactorily operated.

Example. Cable 1

Recently return loss measurements were made on a loaded trunk cable which had a split-gauge makeup and a rather poor loading layout. This cable was placed during 1949 without regard to any special splicing procedures. The cable makeup and other pertinent data are shown in Fig. 1.

Fig. 2 is a block diagram of the testing arrangement and the equipment which was set up at the main frames. The precision balancing and terminating networks were chosen on the basis of the most predominant gauge, and were built out for the best return loss. This usually means building out the far end to a full loading section, and the near or measuring end to a matching section. A Hewlett-Packard oscillator was used as source power for the single-frequency measurements. Echo return loss measurements were made by replacing the oscillator with a Western Electric 201-*A* noise generator equipped with a 455-*B* network which simulates the output of an *F1* transmitter energized by a typical male voice. The *2B* noise-measuring set with *F1A* weighting was used as the indicating instrument. It was found necessary to use the *2B* set because there was insufficient gain available to utilize a vacuum tube voltmeter.

Measurements were made at 500-cycle intervals in the echo band from 500 to 2,500 cycles, and the results for each pair were averaged and plotted as curve *A* of Fig. 3. Curve *B* of Fig. 3 shows the results at 0.707 of the cutoff frequency, and curve *C* shows the results at 1,000 cycles. Curve *D* is the echo return loss as measured with the noise generator.

These four curves were plotted for the purpose of comparison, because design methods and available data variously refer to the 1,000 cycle, the 0.707 cutoff, and the echo return loss. Complete analysis of a cable thus requires measurements of these values, and also at single frequencies throughout the passband of the facility. Fortunately, however, in the usual case it may be assumed that the cable meets the design values, provided that the distribution plotted from measurements at a given frequency compare satisfactorily with the calculated distribution at that frequency.

The return loss distribution will not definitely show that a repeater will work satisfactorily on any individual pair. However, since present methods of measuring return loss on working pairs in some areas require setting up test equipment at the main frames, two men for testing, and others to operate make-busy relays, these tests are expensive. More economical methods usually may be utilized to determine the suitability of individual pairs, once it has been determined that the cable meets the design requirements.

The return loss as shown by Fig. 3 is generally poor, and is about 5 db less than expected. As precision networks are not available for exact matching of split-gauge trunks, some error is likely. It is also possible that a loading coil be missing or that some other trouble exists. No attempt has been made to determine all the factors which may be contributing to the poor return loss of this particular cable, because it was picked at random from a large number of cables in a metropolitan area, and the curves of Fig. 3 were plotted to show the return loss as found, as an illustration of what might be expected of existing cables. The split-gauge makeup may also be partly responsible. Whatever the cause, if the results as shown in Fig. 3 prove to be typical for the older split-gauge cable with poor loading layouts, many existing trunk cables may be unsuitable for economical repeater operation.

EXAMPLE. CABLE 2

This cable was placed in about 750-foot lengths, and each unit or group of pairs within the sheath was random spliced at the points where the sections were joined. The loading layout, with a calculated 0.01 coil-spacing deviation, was considered good. Echo return loss measurements were made with the 201-*A* noise generator method, and the results are shown in Fig. 4.

It may be seen from Fig. 4 that this cable has a very good return loss for a high percentage of the pairs. The measured value for 50% of the pairs is equal to or greater than 34.2 db. This agrees very closely with the calculated value for the 50% point of 34.4 db. Since the measured and calculated values are closely in agreement, it appears reasonable to assume that for this cable the return loss distribution at 0.707 of the cutoff frequency, or at 1,000 cycles, would also meet calculated values. This, of course, may not be true in all cases, as irregularities may cause rather sharp peaks and valleys for different frequencies.

Conclusions

It is possible that many existing loaded trunk cables have poor return losses, and thus must be improved or replaced in order to meet today's transmission objectives. The percentage of pairs having low return loss may be determined by calculation and by measurements for any frequency, or for the echo band, and the return loss may be increased to a satisfactory value by various existing methods.

The return loss distribution for newly placed cables may be plotted from acceptance tests. Any required improvements should be made before placing the cable in service.

References

1. IRREGULARITIES IN LOADED TELEPHONE CIRCUITS, George Crisson. *Bell System Technical Journal*, New York, N. Y., vol. IV, 1925, pp. 561–85.

2. SIMPLIFIED TRANSMISSION ENGINEERING IN EXCHANGE CABLE PLANT DESIGN, L. B. Bogan, K. D. Young. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 498–502.

Discussion

Flay G. Dellinger (Southern Bell Telephone and Telegraph Company, Charlotte, N. C.): The paper by Mr. Speake is a timely one and should stimulate thinking among people who are concerned with the problem of adequate return losses on loaded trunk cables.

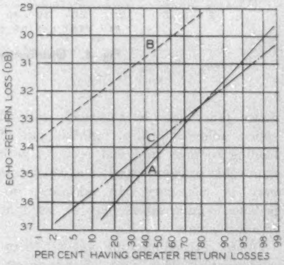


Fig. 5. A—Data from paper by G. H. Speake. Cable 2, 4.5 miles, 19-gauge
B—Trunk cable make-up, Charlotte, N. C. 3.82 miles, 19-gauge HC H-88, 0.90 mile, 22 gauge HC H-88, no capacitance deviation test splicing
C—Trunk cable make-up Charlotte, N. C. 5.14 miles, 22-gauge HC H-88, 437 pairs. No capacitance deviation test splicing

It should be apparent that the design and testing of loaded trunking cables have been handled very loosely in the past. Today, with our via net loss transmission objectives, our so-called exchange loaded trunking cables obviously must meet the same requirements from a return loss standpoint as those of a loaded toll cable. Further, there seem to be wide differences of opinion as to what type of test determines return loss. For example, is it necessary to make both echo return loss tests and singing-point tests? A standard Bell System practice for acceptance tests on trunking cables, I believe, would be most helpful to the Bell Telephone operating companies.

Fig. 5 indicates data taken on trunking cables in the North Carolina area as compared with the Cable 2 results given in Mr. Speake's paper.

G. H. Speake. The author is in full agreement with Mr. Dellinger's comments, which indicate that close attention to design and testing of loaded trunk cables is essential for today's transmission objectives. Also, Fig. 5, submitted by Mr. Dellinger, gives return losses on loaded trunk cables in the North Carolina area which are in general agreement with the author's experience.

A Basis for Transmission Performance Objectives in a Telephone Communication System

W. K. MacADAM
MEMBER AIEE

IN ANY extensive telephone system, certain objectives and controls must be set up to provide and maintain satisfactory performance. This control of quality is as important in communication service as in manufacturing. Rather than stemming from an engineer's dream or a mathematical formula, quality is essentially a user's impression of the product. There is danger, therefore, in measuring it in terms of somewhat arbitrary physical limits. There is danger, too, in not recognizing what lies behind the objectives and the effects on the user when these are not met.

Progress and improved living standards demand continually improving physical performance to maintain the same degree of quality from the customers' viewpoint. In some ways decibels (db) depreciate just like dollars. Hence the importance of not adopting a physical measurement of noise, volume, distortion, for example, and holding it sacred and unchanged throughout the years.

The evaluation of service performance and customer desires has been carried out more or less continuously in the telephone industry since its early days. However, recent subjective tests of received volume have added to the knowledge, particularly with regard to the higher levels. Such information permits the study of reactions to volumes ranging from excessive to inadequate. With this spread in data, the communications product can be analyzed from the viewpoint of preference rather than tolerance. Exposition of the new volume preferences and the derivation of corresponding transmission loss objectives are the main purposes of this paper.

The Service Package

There is no hesitation in naming "service" as the major product of a telephone communication system. Before measuring its quality, it is worthwhile to review briefly what this term includes.

The customer-oriented telephone service package can be divided into four parts, all of which are essential to a complete communications product. Stated in one sentence:

> A customer expects the right features, when he needs them, and as long as he needs them—but always with satisfactory quality.

The Right Features

In the first place, the user must be given the *kind* of service he wants, the type, the color, the shape, the operating features needed in his modern home or business.

When He Needs Them

The second part of the service job is to give the user telephone service *when* he needs it. Not only does this mean giving him a telephone when he wants it, but it also means giving him dial tone promptly when he lifts his receiver. It means that he should get through to his destination without a switching delay, with little chance of receiving an "all trunks busy" tone. In other words, to give service when wanted, not only are sufficient quantities of telephones and voice channels needed at the right places, but also appropriate switching equipment must be present to give him prompt connection.

As Long as He Needs Them

The third component of service is giving the customer service *without interruption*. In other words, this places a premium on continuity of service; freedom from line or equipment failure. To fall short in this respect is to fail to give him a complete service package.

With Satisfactory Quality

The fourth and by no means least of the components of service is the one of primary interest in this paper. This is *transmission*, the quality of the product the customer receives. By this is meant the volume, the fidelity, and the freedom from disturbance that make up a quality communication product.

The first three aspects of the service job give the telephone customer the features he wants and put him in a position where he can carry on communication. Transmission, the fourth side of service, actually is a measure of the product that is finally delivered. With much of the personal contact with operators gone as the result of customer dialing, telephone users judge performance more and more by the quality of the product they receive and less from their reaction to "the voice with the smile." It is significant to realize that, more and more, a customer's main contact with a telephone company is through the receiver of his telephone and the bill he receives every month. Putting both of these together, he will judge the company by what he actually gets for the money he spends. Quality control of the product is a way of seeing that he gets his money's worth.

"Quality of service" is a state of mind of the customer and not any technical measurement. Consequently, this state of mind will change and will be affected by all four of the components noted. As much dissatisfaction can result from a succession of minor annoyances as from a single major service failure. This is why consistently inferior quality is such a subtle danger; why good service one minute and poor the next builds up a sensitivity to contrast. This state of mind is reflected not only in the customer's opinion of the company in its ability to provide a communication channel, but also, rightly or wrongly, in his opinions regarding the company's financial soundness, for example, or whether it is a good place to work, whether it is efficient or a credit to the community.

Although there are a number of transmission effects that make up quality, one of particular interest is volume, since it relates to transmission loss objectives in the telephone plant. For this reason this paper will deal primarily with volume and loss objectives.

Volume by the Foot

As a starting point it is desirable first to obtain some physical idea of how telephone listeners react to various volumes. It is convenient to express relative ease of conversation in terms of an equivalent separation between talkers under more or less reflection-free conditions such as an open field, for example. This device has been used at times in the past.

Table I shows how average ease of conversation has improved over the years for typical telephone calls. It is based on extrapolation of comparative observations made in an open field by engineers from the American Telephone and Telegraph Company and Bell Laboratories in

---

Paper 58-59, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 11, 1957; made available for printing November 7, 1957.

W. K. MacAdam is with the American Telephone and Telegraph Company, New York, N. Y.

1939. The comparative variation with time is probably more accurate than the absolute values. From Table I it may be seen that 35 years ago for intercity calls, customers conversed as though they were 70 or 80 feet apart. For the trunked interlocal calls this "separation" might have been in the neighborhood of 40 or 45 feet. Twenty-five years ago, the average long-distance call dropped to about 35-foot separation with 25 feet for the interlocal calls. In the more recent past, the long-distance service improved to the point that the average toll call was equivalent to conversation at 15 feet with the local call dropping to 10 or 12 feet. Further recent improvement has brought the present-day picture for combined toll and local interoffice calls to an average of about 6 feet. Averages are deceiving, for the spread is quite great as a result of distributions in loop and trunk losses. The average is no consolation to the man receiving poor service on an individual call, particularly if he makes this call to one particular destination every day.

All this shows that quality standards are moving forward. When present performance is measured against listeners' preferences, it is not unusual to find there is still work to be done even if design objectives for new plant are kept up to date. A fixed gun setting is useless on a moving target. The difference between existing plant performance and present objectives is largely the result of not bringing old plant up to date. To some degree it is also caused by failing to get new plant installed and operating according to design.

What the Customer Wants

The first step needed in developing objectives is to find what quality of transmission the customer really wants. What should be the basis for service aims: complaints, tolerance, general satisfaction, or preference? Is there a demand for good service even if it costs somewhat more than poor service? Will poor service be tolerated under anything but emergency conditions? More fundamentally still, one is faced with the question: "What is 'good' and what is 'poor'?"

Table I. How Bringing Customers Closer Together Is a Continuing Job

When	Distance in Air, Feet: Intercity	Distance in Air, Feet: Trunked Local
35 years ago	70–80	40–45
25 years ago	35	25
Recently	15	10–12
Now	6 feet average	

Notes:

1. Weighted for bandwidth, sidetone, etc.
2. Individual calls may differ considerably from values shown.

Operating from complaints alone would emphasize objectives based only on what the customer will tolerate. This is like "Excellent" ratings comes at higher volumes (or lower circuit losses) than the valley for "Poor" ratings. This is caused by fewer "Good" (loud) ratings compared with "Good" (low). People seem to be so happy with high volumes that they rate them "Excellent" up to the point that their ears tingle and they begin to drop the rating into the "Poor" range or worse. At slightly lower volumes there is a larger range where many of the calls are either "Excellent," "Good," or "Fair," but very few "Poor."

Here is where a little judgment must be exercised to consider the economy and the effect on the telephone bill. At the high volume levels, at the peak of the "Excellent" ratings which were made without cost consideration, the user is fairly luxuriating in volume. He figuratively has his furnace on full, his windows open, and the gas meter by-passed. It is necessary to come down to earth at this point and look for prudent comfort and not princely elegance. This means swinging sights away from the peak of "Excellent" toward "Good" (low volume) but keeping "Poor" at a minimum.

There is another factor that must be kept in mind in setting design sights. Over-all service objectives must include the combined effect of design plus occasional troubles which produce poor volume or circuit failure. Telephone users do not worry as to whether their volume difficulties were caused by design weakness, component failure, poor adjustment, or a wet cable. They are just as unhappy in any case. Hope is always present that the millennium will come, that all plant will become trouble-free and will be kept in perfect adjustment. Even with all the attention engineering and maintenance people are giving this problem, these sources of low or nonexistent volume will continue for some time. Perfection is not necessarily the best over-all economic solution here any more than in the case of design. Consequently, to make due allowances for these sources of difficulty, even tighter reins must be placed on average design objectives to keep the over-all service in check. Allowing for an occasional missed target is similar to the philosophy of the taxi driver who said, "Twenty-five cents is my average tip but there ain't very many that gives the average."

With these considerations in mind, and recognizing that, for short hauls, volume still costs the telephone company and its customers something, design objectives might be set something like this:

Negligible number of calls rated *poor* because of design.

No more than about 5% rated *fair.*

The balance, *good* or *excellent.*

From Fig. 1 it is seen that this corresponds to the line marked "present objective" with an average trunk loss distribution of about 4 db. This is the current goal in the Bell System for trunked calls. To obtain this average of 4 db, total trunk losses should be controlled so that they range, say, between 2 and 6 db, assuming typical loop loss conditions.

In the future, when carrier becomes economical for shorter lengths, and impedances are under better control higher volumes will be available at practically no additional cost and further improvement will be possible. Nevertheless, it is equally certain that the future will bring preferences for still higher volumes. Probably more important still will be a trend toward lower talker volumes as the users become accustomed to the feeling of nearness to the other party. All this means future requirements for still lower trunk losses. In the ultimate, the objective must approach zero loss. From there on it must be lower loop loss, for the following reason.

Loss Reduction: Loops Versus Trunks

Controlling volumes means essentially, controlling the over-all loss between lips and ear. It is important, however, to apportion it properly over the various types of plant. This cannot be done just any way. The allocation must consider the effect on service, the relative usage of loops and trunks, and the cost and quantities involved, with, at the same time, future possibilities in sight. As accident prevention based only on a count of mortalities. Actually, to give satisfactory service in all respects, it is important to know what the user prefers and what he needs, and not just what he will barely stand.

Subjective tests by Reier and Coolidge at the Bell Telephone Laboratories have developed considerable data on received volume preferences in telephone conversations. For a given volume, several thousand observations were made with a variety of typical listeners. Volumes were rated by these listeners in several

Vol.
Pt. 1
1958

categories such as "Excellent," "Good," "Fair," "Poor," etc. Both appraisal and comparison tests were made in the range between "Intolerably Loud" and "Could Not Understand." From these data a family of curves was derived showing, for any particular received volume, the per cent of listeners that would rate this volume as "Good," "Excellent," etc.

Recently the Reier and Coolidge data were used as a basis for evaluating listener preferences under various circuit loss conditions. This was done by determining the distribution of received volumes which results from combining typical talker volume distributions with various trunk and loop loss distributions. Fig. 1 is the result of this statistical combination, considering trunked calls of any type. It shows how these trunked calls would be rated by a typical variety of customers for clusters of trunk circuit losses having different average values. The relationship between effective loop loss and volume loss is not a straightforward adjustment since it depends on frequency reponse, noises, and other factors. Although this relationship is still under study, the curves were developed using Reier's data of several years past.

Since Fig. 1 is used primarily to study trunk losses, the preferences are plotted in terms of various trunk loss averages for a constant typical cluster of loop losses. This assumes a fixed loop loss distribution with spread and average such as frequently encountered. If the average loop losses differ from those assumed, a compensating (−db or +db) change must be made in the trunk loss averages read on the horizontal scale. Care must be taken not to consider the curves as typical probability curves since the horizontal scale is in terms of averages and not specific individual losses.

Setting the Thermostat

Equipped with information on the customer's volume preferences, and some idea of what he thinks quality is worth, the next job is to set design objectives. This is not a matter of blindly increasing the gain. Adjusting volumes on telephone circuits is somewhat like setting the home thermostat in cold weather. Volumes can be too high or low for comfort, just as in the case of home temperature. On the average, high volumes are more costly, just as is the fuel bill for a high thermostat setting. Of course, where carrier proves in, this added volume cost is slight.

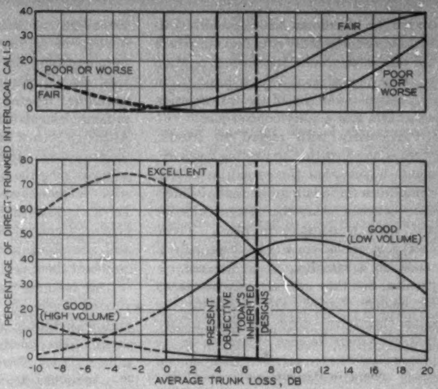
Increased temperature brings increased comfort only when the house is cold.

Fig. 1. Customer rating of transmission; trunked calls

Assuming standard deviation of trunks = 2 db

Assuming standard deviation of loops = 1.8 db

Assuming average of loops = −4 db effective



Likewise, higher volumes mean greater satisfaction only when normal levels are on the low side. At times in the volume case, not only is the "temperature" low, but past designs have set the thermostat below the present-day comfort level.

With reference to temperature preferences, some want it cool while others prefer it warm. Because of this spread, combined with the up and down variations in the heating system itself, there is only a comparatively narrow range left for the setting before someone becomes dissatisfied. The same is true for volumes. For any specific average trunk loss, some calls will be made over short trunks or loops to people who are sensitive to loud volumes. Likewise, there will be some calls which encounter more than average loss and will involve people who dislike low volumes. The net result is a very large spread in ratings on all calls for any particular design objective. Design averages cannot shift very far up or down the scale before some listeners become dissatisfied. There is not much of a range in which to operate.

Even in this relatively narrow design range it is wise to set the volumes on the low side of optimum. Experience shows that users will not be satisfied with bad service even at a bargain, so economic considerations lead to an objective somewhere between what is just tolerated and what in itself is preferred, but biased in the preferred direction. This calls for study of Fig. 1, recognizing that the preference tests were made in the absence of cost consideration. The first thing probably noticed is that the peak of the previously mentioned, Fig. 1 was drawn up for trunked calls, assuming a constant loop loss distribution with varying average trunk losses. It could, if necessary, have been made up on an over-all circuit loss basis or could have assumed a constant average trunk loss and a variety of loop loss averages. This was not done because at present about all the loss possible has been squeezed out of loops without getting into big money.

There are many ways to divide the circuit to apportion the loss objectives for trunked calls. Loops and trunks, however, are the natural dividing points, principally because the usage on an individual loop is so much smaller than that on a trunk shared by many users. On an average, a trunk is used about five times as often as a loop. This, coupled with the fact that loops outnumber trunks more than ten to one, makes one willing to spend more money per circuit on trunks than on loops to help transmission on trunked calls. It is simply a matter of getting more for the money.

BROAD OBJECTIVES

This generalized philosophy calls for first doing the things that are economical and that help control loop losses: "For loops, get the best set efficiency, use the high gain sets where needed, use loading where needed, use fine gage to the extent one can signal over it, and then put the burden of loss reduction on the more efficiently used trunks." Before much more can be done to the loop design, further progress is necessary in cutting trunk losses.

Discovering What Needs to Be Done

When objectives have been derived from what the customer himself believes necessary for satisfactory service, the actual job ahead can be measured. The first step undoubtedly is to find out where the telephone system stands with relation to these objectives. This is basic to an evaluation of the size of the transmission correction and quality control job.

Particularly with regard to trunks, there is no substitute to making measurements to see what the actual losses are. Sometimes there is a great deal of difference between what the design calls for and what is actually in place. These checks of actual conditions, especially if made on a sampling basis, are not expensive, but they are certainly necessary for quality control. Of course, the designs themselves need checking to see that they meet modern objectives or are programmed for correction. For loops a check of designs and proper use of the more efficient telephone sets on high-loss loops is also desirable.

Although set surveys and trunk measurements are important, they are only indirect measurements of what the telephone user feels about the quality of service. The closer one gets to the customer in examining the quality of service, the better is the measure of the job being done and of the work that lies ahead.

Three principal schemes are currently being used in the Bell System to measure the product as it is actually delivered. In the first place, a selected sampling of customer opinion is presently being made annually by means of home visits. Known as the "Customer Attitude Survey," it includes questions on the cost of service, quality of service, and opinion of the telephone company. The "quality of service" portion of this survey contains questions on the grade of transmission experienced. The results are published in a booklet entitled "What Customers Think" for information of all departments. By analysis of these periodic surveys, information can be obtained as to the transmission health of the various areas and the locations where further study or surveys need to be made in order to pinpoint the sources of trouble.

The particular questions involving transmission are

If *out-of-town* calls are made, is there anything about the service on these calls that is in any way unsatisfactory?

Possible answer: Yes, or No
(If yes) What is unsatisfactory?.......

When using your telephone at home do you ever have any trouble hearing or being heard?

Possible answers: Very often
Fairly often
Only occasionally
Almost never

In addition to direct customer surveys, two additional means are used to measure quality of service. These are volume measurements on trunk appearances in local central offices, and observations of customer repetition rates and transmission irregularities on intercity calls. Although both of these emphasize physical measurements, they have the advantage of measuring the quality of calls in progress, rather, than indirectly evaluating service by idle trunk measurements. The latter, of course, serve a useful purpose and the direct quality surveys will not supplant them.

The volume survey program was started early in 1957. A high-gain volume indicator is used and a record is made of the volume of the more distant talker. The distributions of volumes measured at each central office are related to the established listener preferences of "Excellent," "Good," or "Poor," adjusted for the loss between the central office and subset. The present goal is to survey volumes in each central office about once a year. Initial results of the volume measurements show about 4% "Poor," 15% "Fair," and the balance "Good" or "Excellent," somewhat short of the current goal. The toll repetition rates are noted on a sample of calls at service observing positions, with a count being made of the number of repetitions per 100 seconds. Note is also made of calls in which customers comment adversely on quality or where, in the observer's opinion, noise, crosstalk, or other impairments are excessive.

The intercity service observations have been particularly useful in comparing performance between different localities and in studying trends. Work is now under way in an attempt to relate the repetition rate performance to volume, and hence to user ratings of "Good," "Fair," etc. It is also proposed to study the advantages of volume measurements on toll connecting trunks to augment the service observations.

Considering trunk design itself, it is recognized that work is needed to bring plant up to the current goal. Present-day trunks are a combination of fairly recent additions and past construction designed for former lower volume objectives. A survey made in 1956 in the Bell System shows that the resultant existing designs, both new and "inherited," range to 12 db or higher. It is reasonable to believe that similar conditions exist in other telephone systems. The average is about 7 db and corresponds to the second dashed line in Fig. 1 marked "Today's Inherited Designs." The ratings for this average loss are in fair agreement with the results of the volume surveys, recognizing the difference between designs and actuality.

Tandem trunk designs in general are still running 2 to 7 db with an average of about 4 db. Since two tandem trunks are required for a tandem call, this results in about the same total trunk loss as the direct path. Note once more that this refers to design and not necessarily to what is actually in service.

From Fig. 1 again, the measure of the job ahead can be seen. Almost 2% of all trunk designs are in the "Poor" category with 11% rated "Fair," and the balance "Good" or "Excellent." Based on current estimates of trunked calls, this means that each day customers would rate about 1.5 million calls "Poor," assuming that performance and design were identical. Fortunately the importance of improved volume is generally being recognized and programs are under way in many locations to bring about progressive improvement.

Conclusions

The work with customer preferences in the field of telephone transmission must form the backbone of design objectives. It demonstrates the inadequacy of cradle-to-the-grave limits in a world of changing preferences.

An appreciation of the background of customer-derived objectives permits establishment of an intelligent and prudent program for up-dating existing plant to modern requirements. It emphasizes the flexibility needed in present designs to care for future changes. With service the main telephone product, its quality control is not a luxury but a business necessity.

References

1. TRANSMISSION ENGINEERING FOR THE FUTURE, H. R. Huntley. American Telephone and Telegraph Company, New York, N. Y., 1955.

2. A REVISED TELEPHONE TRANSMISSION RATING PLAN, P. W. Blye, O. H. Coolidge, H. R. Huntley. *Bell System Technical Journal*, New York, N. Y., vol. 34, May 1955, pp. 453–72.

3. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, H. R. Huntley. *Ibid.*, vol. 32, Sept. 1953, pp. 1019–36.

◆

Discussion

J. D. Howard, Jr., and J. W. Joyner (Southern Bell Telephone and Telegraph Company, Atlanta, Ga.): This excellent paper describes in simple language the present service goals of the industry and their derivations. If we conscientiously design to these objectives and keep them current and ever before us in the operation of our plant, a good public opinion and its rewards will be ours.

Our current measurement program shows that we are quite a way from the 95% "good," 5% "fair" which we accept as the "happy medium" operating point. In Southern Bell, only 80% of our calls were considered "good" by the customers for 1957. This is probably due as much to a general lack of awareness as to a poor plant inheritance.

Unless the installer who puts in the telephone, the test-deskman who checks the service, the outside plant engineer who designs the loop, and the multitude of others involved in providing telephone service are all aware of the part that their specialty plays in shaping subscriber opinion, the most sincere transmission design is for naught.

Improvement in the quality of telephone service (speed of connection and ease of conversing over the connection) has improved far more than the increase in the cost of that service to our customers and this is as it should be; but if our quality is to continue to improve, we must obtain rates that enable us to program our corrective measures and engineer our new plant to meet our customers' wishes consistent with sound economy.

R. C. Crawshaw (New York Telephone Company, Syracuse, N. Y.): I think Mr. MacAdam's paper spells out a "modern" approach to the transmission job for telephone engineers. It brings the job down to earth. The very end result of many technical endeavors is usually a product to be used by the public at large, whether it is a refrigerator, an automobile, or a telephone system. The engineer, although he must have technical know-how and must deal and think in specialized terms in developing the product, must ultimately have his efforts measured in terms of not what he, the developer, thinks of the technical soundness of the product, but what the user thinks of it as a thing well worth the money. It may be the former (the technical approach) that produces the product but it is the latter (the customer's opinion) that sells it.

Some transmission engineers of the past, and I fear too many, have lived in a kind of ivory tower. They not only set up their own standards to measure the product of their thinking but, what is worse, try always to explain objectives to the uninitiated in technical terms not fully understood by their audience. To use a homely example, this could be like the law breaker setting up his own standards for the measurement of his own innocence (or transgression?). The proposed quality objectives avoid this trap.

This paper performs a very useful task in combining a down-to-earth basis of viewing the effectiveness and real worth of the basic product we sell—transmission—in terms of the customer's opinion of how well it serves his needs and uses. It also defines a better way of explaining it so that transmission engineers can more easily establish their priority for money needs built into construction programs or company budget setups. These remarks are meant to emphasize the need for transmission engineers to make this approach a way of clearly defining the rightful balance between money requirements for "switching" and for "transmission." Otherwise we may find that we have spent millions on far-reaching switching machinery that switches—what?. We may find it all just a technical dream, and perhaps one with which the person paying the bill is not satisfied. Only by measuring the quality of our transmission product both service- and cost-wise from the bill payer's angle can we really know what our mission is and how well we fulfill it. The more our world becomes an engineer's world, the more we need to recognize the need for engineers to be not only more adept technically, but also more adept at interpreting and measuring their field of endeavor in terms of actual human interest and usefullness. This quality control concept of transmission brings our communication system in line with this objective.

Although not mentioned as such in the paper, it is implicitly implied that we must not only place weight on the design and installation of our product, but on the maintenance of it within customer-accepted standards since this too is an essential to quality control. To do this we should know the standards from the customer's view and three ways of measuring our status in this respect are cited. However, I believe there is an interrelation between design and these measured values that does not receive enough attention. It is a need if we are to meet the objectives set up by Mr. MacAdam, and so it is well at least to mention the practical side of keeping our actual performance in line with objectives as we go deeper into automation, with loss of employee contact, and as we work down to lower losses.

First, we must be sure that we design into our plant and equipment proper maintenance margins. I do not mean margins as an allowance for poor maintenance, but with a margin to strike the proper balance between investment and maintenance costs. The more strictly we maintain design tolerances, the more likely it is that the maintenance job will become more technical than the "going" know how of the talent we *can afford* to assign to the everyday job, and perhaps than we can produce by realistic hiring and training procedures. It is a matter of economics. I would suggest some soul-searching on the part of design and circuit layout engineers to be sure we get the right balance between investment and maintenance costs. We should be sure that we are not relying too much on the maintenance side of the ledger in trying to meet our customer objectives.

Second, some further thought might well be applied to the test gear we engineer for the maintenance man. These things too are an important part of the engineer's job in quality control and in meeting the customers' needs economically. One has only to look at a list of the growing amount of test gear applied to all types of transmission channels these days and the cost and effort required to keep them calibrated and working right to realize there may be something to this point in furthering our transmission objectives from the customer's point of view.

I suspect these two points sum up the basic reasons that have tended to bring the transmission maintenance engineer again more prominently into our thinking in recent years.

The measurement of the quality of our product from the customer's angle is a realistic guide in setting technical standards of design and maintenance. Further work along this line is needed as is pointed out in the paper. The control of this quality is in the hands of the design and maintenance engineer as far as providing the gear and the tools for its accomplishment. Our programs need to be continuously scrutinized to be sure we are meeting these standards with the proper balance between investment and maintenance dollars.

W. K. MacAdam: I was much interested in the thoughts and constructive comments of the discussers.

It seems to me that Mr. Howard and Mr. Joyner have put their finger on a tremendously important aspect of providing good telephone service, or for that matter, any other kind of customer service: compensatory rates. Only a company that is in good shape dollar-wise can stay in good shape equipment-wise; and the provision of modern equipment, properly and adequately maintained, is essential to the provision of the quality of service customers prefer. If one needs more evidence in support of this contention he might ask the next suburban commuter he meets in virtually any large city.

Mr. Crawshaw's remarks suggest that a large part of the burden for effective maintenance rests squarely on the shoulders of the engineers designing and planning for new facilities. This is certainly true. In fact, the problem is becoming more acute daily as more and more electronic devices find their way into the telephone plant. We must not only be increasingly diligent but we must learn how to be more ingenious. When engineers recognize that they share the responsibility for good maintenance, however, I am sure that we are off to a good beginning.

An Electric Circuit Theory Approach to Finite Difference Stability

WALTER J. KARPLUS
NONMEMBER AIEE

ONE of the major difficulties involved in the numerical solution of partial differential equations of the parabolic or hyperbolic type by means of finite difference approximations is the tendency of the solution to be unstable under certain conditions. If instability exists, a small error (e.g., round-off error) arising at some point in the computation procedure, tends to become larger and larger as the computation progresses, until the error terms completely overshadow the desired solution, making it worthless. Some finite difference approximations appear to be stable under any condition; others are always unstable; while some are stable only if the spacing intervals satisfy certain requirements.

The two most familiar criteria for testing difference equations for stability are those due to Hildebrand[1] and to von Neumann as described by O'Brien.[2] In both of these methods, it is assumed that the solution of the finite difference equation may be expressed as a product of functions, each depending on only one of the independent variables of the equation. This general solution is then analyzed to determine what requirements must be met by the coefficients of the equation, so that a computational error necessarily decays when one of the independent variables (usually time) goes to infinity.

A different approach to the stability problem follows from the recognition that the voltage or current distribution of a network of electrical resistors arranged in a regular pattern may be expressed compactly by a finite difference equation. Conversely, a finite difference equation may be considered to have an electrical resistance network analog. It follows that any instability in the finite difference network must be reflected as electrical instability in the analogous electric network. Where a small computational error in a numerical finite difference computational procedure gradually tends to grow to infinity, so in the electric network the appearance of a small voltage or current "noise" may be expected to grow indefinitely. This paper represents an effort to apply circuit theory concepts, for dealing with electrical instability, to the problem of determining whether a given finite difference equation, representing an approximation to a partial differential equation, is stable or unstable.

Stability of Resistance Networks

A homogeneous rectangular resistance network may be thought of as being composed of "building blocks" of either of the two basic types shown in Fig. 1, which are repeated in both space directions (*m* and *n*) either to infinity or until a specified boundary is reached. Kirchhoff's current law $\Sigma i_n=0$ relates the potentials V of the adjacent nodes to the potential at node *m*, *n*, in Fig. 1(A):

$$\frac{V_{m-1,n}-V_{m,n}}{R_1}+\frac{V_{m,n+1}-V_{m,n}}{R_2}+\frac{V_{m+1,n}-V_{m,n}}{R_3}+\frac{V_{m,n-1}-V_{m,n}}{R_4}=0 \quad (1)$$

while Kirchhoff's voltage law $\Sigma e_l=0$ expresses the loop current of loop *m,n* in terms of the currents in the adjacent loops:

$$R_1(i_{m,n+1}-i_{m,n})+R_2(i_{m+1,n}-i_{m,n})+R_3(i_{m,n-1}-i_{m,n})+R_4(i_{m-1,n}-i_{m,n})=0 \quad (2)$$

The resistance network is considered stable if a perturbation voltage at node *m,n* in Fig. 1(A) or a current in loop *m,n* of Fig. 1(B), which may be due to noise or other causes, approaches zero and dies out after excitation is removed. Evidently if all resistors have positive values, the network must be stable, since it then contains only dissipative elements and will return to an electrically quiescent state as soon as all the applied energy has been dissipated. It is the presence of active elements in the form of negative resistors which may or may not cause the network to feed back more energy to the node or loop under consideration than was applied originally.

A complete analysis of a large network, containing negative resistors to determine its response to a small transient excitation at some loop or node within the network, can be laborious. A more convenient approach is to recognize the regular repetitive nature of the network and to reason that if a typical building block of the network is stable, i.e., if the building block acts to dissipate rather than supply energy in response to an excitation, then the entire network must be stable. This must be true, since any energy not dissipated by the block to which the excitation is applied will be fed to adjacent blocks which are also dissipative in nature. The problem is reduced therefore to the specification of the conditions under which a typical building block is stable. In the current analog, Fig. 1(A), this implies grounding all nodes except node *m,n* and applying a voltage impulse to this node. If the network is stable, no current will flow through the resistors after the excitation becomes zero. Likewise, in the voltage analog, Fig. 1(B), all loops except loop *m,n* are opened, and a current source is inserted in the loop. If the network is unstable, a current will continue to flow through the loop after the current source is removed.

To determine under what conditions typical network units, as shown in Fig. 1, are stable, it is necessary to consider in detail the operation of negative resistors in electrical circuits. Such a discussion is presented in the Appendix. All positive resistors, comprising the typical node or loop, are lumped together into a single resistor of magnitude R_p, while all negative resistors are combined to form a single negative resistor of magnitude R_n, as shown in Fig. 2. It is then demonstrated that the typical node Fig. 2(A) is stable only if $R_p>R_n$, and that the typical loop, Fig. 2(B), is stable only if $R_p<R_n$.

Application to Finite Difference Equations

The application of the foregoing theory to finite difference equations is best explained by considering, as an example, the wave equation

$$\frac{\partial^2 w}{\partial x^2}=\frac{\partial^2 w}{\partial t^2} \quad (3)$$

and its customary difference form

$$w_{m+1,n}-2w_{m,n}+w_{m-1,n}=k^2(w_{m,n+1}-2w_{m,n}+w_{m,n-1}) \quad (4)$$

where $k\equiv\frac{\Delta x}{\Delta t}$

$$w_{m,n}\equiv w(x,\ t)$$

$$w_{m+1,n+1}\equiv w(x+\Delta x,\ t+\Delta t)$$

Paper 58-60, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 10, 1957; made available for printing December 9, 1957.

WALTER J. KARPLUS is with the University of California, Los Angeles, Calif.

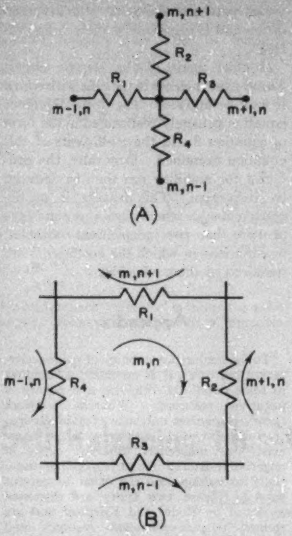


Fig. 1. **Building blocks of resistance network**

A—Node basis B—Loop basis

This expression, when rearranged as

$$(w_{m+1,n}-w_{m,n})+(w_{m-1,n}-w_{m,n})-k^2(w_{m,n+1}-w_{m,n})-k^2(w_{m,n-1}-w_{m,n})=0 \tag{5}$$

is seen to be of the same form as equations 1 and 2. In accordance with the latter of these equations, equation 5 may be identified with a resistance network, whose basic building block is a loop of four resistors: $+1$, $+1$, $-k^2$, and $-k^2$ in magnitude. The stability condition that $R_p<R_n$ then implies that the finite difference equation 4 is stable only if

$$(+1\ +1)<(+k^2+k^2) \tag{6}$$

or

$$k>1 \tag{7}$$

This, incidentally, is precisely the same result as that obtained by applying either Hildebrand's or von Neumann's criterion.

The general application of the circuit theory approach to finite difference stability may be summarized as follows:

1. The finite difference equation is arranged in the form

$$1(w_{m+1,n}-w_{m,n})+a(w_{m-1,n}-w_{m,n})+b(w_{m+2,n}-w_{m,n})\ldots+c(w_{m,n+1}-w_{m,n})+d(w_{m,n-1}-w_{m,n})+e(w_{m,n+2}-w_{m,n})\ldots=0 \tag{8}$$

where the subscript m refers to a bounded space co-ordinate, and the coefficients 1, a, b, c, d, etc., are examined.

2. If all the coefficients are positive, the equation is stable.

3. If some of the coefficients are negative, a sufficient condition for stability is that the algebraic sum of all the coefficients be negative.

If the finite difference equation represents an approximation to a partial differential equation, it is in general possible to arrange it in the form of equation 8.

As a second example consider the familiar 9-point expansion of equation 3:

$$k^2(w_{m,n+1}-2w_{m,n}+w_{m,n-1}) = \begin{cases} a(w_{m+1,n-1}-2w_{m,n-1}+w_{m-1,n-1})+ \\ (1-2a)(w_{m+1,n}-2w_{m,n}+w_{m-1,n})+ \\ a(w_{m+1,n+1}-2w_{m,n+1}+w_{m-1,n+1}) \end{cases} \tag{9}$$

which may be expressed as

$$(w_{m+1,n-1}-w_{m,n})+(w_{m-1,n-1}-w_{m,n})-2(w_{m,n-1}-w_{m,n})+\left(\frac{1-2a}{a}\right)(w_{m+1,n}-w_{m,n})+\left(\frac{1-2a}{a}\right)(w_{m-1,n}-w_{m,n})+(w_{m+1,n+1}-w_{m,n})+(w_{m-1,n+1}-w_{m,n})-2(w_{m,n+1}-w_{m,n})-\frac{k^2}{a}(w_{m,n+1}-w_{m,n})-\frac{k^2}{a}(w_{m,n-1}-w_{m,n})=0 \tag{10}$$

The sum of the coefficients of equation 10 is

$$\left(-\frac{2k^2}{a}+2\left(\frac{1-2a}{a}\right)\right)$$

The sufficient condition for stability, therefore, is that

$$k^2>(1-2a) \tag{11}$$

for $a<1/2$, while the expression is always stable if $a>1/2$. This result is slightly more conservative than the corresponding von Neumann condition

$$k^2>(1-4a)$$

The analysis of equations with more than two independent variables is equally straightforward. Consider the 2-dimensional wave equation

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}=\frac{\partial^2 w}{\partial t^2} \tag{12}$$

and its finite difference expansion

$$\frac{w_{m+1,n,\rho}-2w_{m,n,\rho}+w_{m-1,m,n,\rho}}{\Delta x^2}+\frac{w_{m,n+1,\rho}-2w_{m,n,\rho}+w_{m,n-1,\rho}}{\Delta y^2}=\frac{w_{m,n,\rho+1}-2w_{m,n,\rho}+w_{m,n,\rho-1}}{\Delta t^2} \tag{13}$$

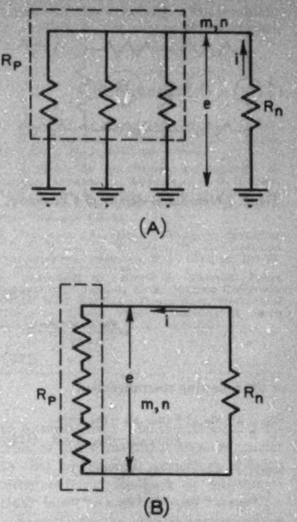


Fig. 2. **Building blocks including negative elements**

which may be expressed as

$$(w_{m+1,n,\rho}-w_{m,n,\rho})+(w_{m-1,n,\rho}-w_{m,n,\rho})+l^2(w_{m,n+1,\rho}-w_{m,n,\rho})+l^2(w_{m,n-1,\rho}-w_{m,n,\rho})-k^2(w_{m,n,\rho+1}-w_{m,n,\rho})-k^2(w_{m,n,\rho-1}-w_{m,n,\rho})=0 \tag{14}$$

where

$$l\equiv\frac{\Delta x}{\Delta y}$$

$$k\equiv\frac{\Delta x}{\Delta t}$$

Summation of the coefficients immediately leads to the condition

$$k^2>(1+l^2) \tag{15}$$

as being sufficient to guarantee stability.

Nor is this approach limited to equations which correspond directly to complete electric networks. With some circumspection it may be extended to equations for which topological considerations permit the construction of an analog of no more than the basic building block. For example, the diffusion equation

$$\frac{\partial^2 w}{\partial x^2}=\frac{\partial w}{\partial t} \tag{16}$$

may be expanded in any of the following ways:

$$\frac{w_{m+1,n}-2w_{m,n}+w_{m-1,n}}{\Delta x^2}=\frac{w_{m,n}-w_{m,n-1}}{\Delta t} \tag{17A}$$

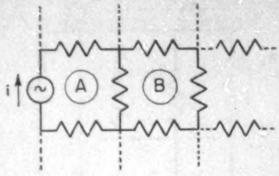


Fig. 3. Network in vicinity of a boundary

$$=\frac{w_{m,n+1}-w_{m,n-1}}{2\Delta t} \tag{17B}$$

$$=\frac{w_{m,n+1}-w_{m,n}}{\Delta t} \tag{17C}$$

or the following rearrangement

$$(w_{m+1,n}-w_{m,n})+(w_{m-1,n}-w_{m,n})+k(w_{m,n-1}-w_{m,n})=0 \tag{18A}$$

$$(w_{m+1,n}-w_{m,n})+(w_{m-1,n}-w_{m,n})-\frac{k}{2}(w_{m,n+1}-w_{m,n})+\frac{k}{2}(w_{m,n-1}-w_{m,n})=0 \tag{18B}$$

$$(w_{m+1,n}-w_{m,n})+(w_{m-1,n}-w_{m,n})-k(w_{m,n+1}-w_{m,n})=0 \tag{18C}$$

where

$$k\equiv\frac{\Delta x^2}{\Delta t}$$

By inspection, equation 18(A) is seen to be always stable, 18(B) always unstable; and 18(C) is stable for $k>2$.

The more complicated parabolic equation

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial w}{\partial x}=\frac{\partial w}{\partial t} \tag{19}$$

is expanded as

$$\frac{w_{m+1,n}-2w_{m,n}+w_{m-1,n}}{\Delta x^2}+\frac{w_{m+1,n}-w_{m,n}}{\Delta x}=\frac{w_{m,n+1}-w_{m,n}}{\Delta t} \tag{20}$$

$$(w_{m+1,n}-w_{m,n})+(w_{m-1,n}-w_{m,n})+h(w_{m+1,n}-w_{m,n})-k(w_{m,n+1}-w_{m,n})=0$$

where

$$h\equiv\Delta x,\ k\equiv\frac{\Delta x^2}{\Delta t} \tag{21}$$

The simple addition of the coefficients of equation 21 then leads to the stability condition

$$k>2+h \tag{22}$$

identical to that obtained by von Neumann's method.

Effect of Boundary and Initial Conditions

Both von Neumann's and Hildebrand's criteria are conservative for any particular problem because they are applicable for any set of "proper" boundary and initial conditions. For a secific boundary condition the actual stability requirement may be less severe. A similar limitation applies to the method just described. An equipotential boundary of a field described by a difference equation is reflected in the electrical analog, constructed on the voltage law basis, as a constant current boundary, (since loop currents correspond to the potential at the node points of the finite difference grid). That is, every loop adjacent to the boundary must include a constant current source, as shown in Fig. 3. It is evident that loop A in Fig. 3 must be stable regardless of the magnitudes of the positive and negative resistors comprising this loop. This is true because the constant current source forces the current in loop A to have a constant magnitude. This forcing effect is felt, to a somewhat lesser extent, in the adjacent loop B, so that the stability condition of this loop is less severe than would ordinarily be the case. As the distance of the loop under consideration from a specified boundary increases, the stabilizing effect of that boundary decreases. Where one of the independent variables, e.g., time, goes to infinity, this effect eventually becomes negligible and the foregoing analysis applies.

In general, it is to be expected therefore that the stability condition derived from an electric circuit approach will be overconservative unless specific account is taken of the effect of the boundary potentials. Occasionally, where a relatively large net interval is employed, it is possible to construct the entire analog network and deduce the stability condition for each region in the network.

Conclusions

A criterion for the stability of finite difference equations, representing expansions of partial differential equations, has been developed. This criterion is based on the stability analysis of an electric network whose equilibrium equations have the same form as the finite difference equation under study. The results of applying this criterion to a number of difference equations when compared with the results of applying the more familiar stability criteria of Hildebrand and von Nuemann, show that the electrical analog approach frequently yields precisely the same stability condition and is occasionally more conservative.

A chief merit of the electric circuit theory approach lies in the ease with which it may be applied. The finite difference equation is merely rearranged in the form of equation 8 and the coefficients of this equation examined. Generally, the condition for stability can then be deduced by inspection. This feature is particularly valuable when examining equations of more than two independent variables, or equations in which the coefficients are functions of these variables.

Appendix

To determine the stability of nonreactive, active networks, it is necessary to analyze in some detail the behavior and nature of negative resistors. Whereas network theory recognizes only one type of dissipative element, the positive resistor, there exist two inherently different types of negative resistors: one corresponding essentially to voltage and the other to current sources. These two types are discussed in detail by Bode[3] and Karplus[4] and are termed current-controlled resistors and voltage-controlled resistors. In a voltage-controlled negative resistor, the current through the element is a specified function of the applied voltage, as determined by the magnitude of the resistor. However, the voltage drop, which is produced as a result of an applied current, bears no meaningful relationship to the magnitude of this current. In this respect, this type of element is similar to a constant current source. Likewise, in analyzing a current controlled device, it is meaningful to talk only of the voltage response to an applied current. Therefore, before it is possible to determine the behavior of this element in an electrical network, it is necessary to know with which type of negative resistor one is dealing.

This consideration becomes important when Kirchhoff's laws are applied to a network containing negative resistors. Since the basic node and loop equations 1 and 2 make no distinction between the two types of negative resistors, it follows that some assumption is made implicitly in writing these equations. For example, Kirchhoff's current law, when applied to a node formed by three positive resistors and one negative resistor, states that, given the current through the three positive resistors, the current through the negative element is automatically determined. The voltage drop across this resistor then adjusts accordingly. The application of the current law therefore predicates that all negative resistors are of the current-controlled type. Similar considerations indicate that the voltage law is applicable only to negative resistors of the voltage-controlled type.

If one of the resistors in Fig. 1(A) is negative, and if equation 1 is to be applicable, this resistor must be assumed to be of the current-controlled type. The stability analysis of the network block then proceeds as follows: all node points except point m,n are assumed grounded; all positive resistors may then be lumped together and designated by R_p, while the negative resistor is designated by R_n as shown in Fig. 2(A). If now a small voltage e appears at node m,n, a current i is produced through R_p, so that

$$i=\frac{e}{R_p} \tag{23}$$

which must return through the current controlled resistor R_n. This current produces a regenerative feedback voltage e' at node m,n

$$e'=e\left(\frac{R_n}{R_p}\right) \tag{24}$$

If this feedback voltage e' is larger than the applied voltage e, the circuit is unstable and the node voltage will rapidly increase to infinity in response to a finite excitation. Stability is therefore possible only if $R_p > R_n$.

Likewise, if one of the resistors in Fig. 1(B) is negative and equation 2 employed, this resistor must be of the voltage-controlled type. All the positive resistors may then be lumped together as shown in Fig. 2(B) and a small current i may be assumed to arise in the loop, all other loops being assumed open-circuited. This current then produces across R_p a voltage e

$$e=iR_p \tag{25}$$

This voltage also appears across R_n and a feedback current i' flowing in the same direction as i

$$i'=i\left(\frac{R_p}{R_n}\right) \tag{26}$$

results. If i' is larger than i, the loop current will grow to infinity. The condition for stability in this case is therefore that $R_p < R_n$.

References

1. METHODS OF APPLIED MATHEMATICS (book), F. B. Hildebrand. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1942, pp. 328–45.

2. A STUDY OF THE NUMERICAL SOLUTION OF PARTIAL DIFFERENTIAL EQUATIONS, G. G. O'Brien, M. S. Hyman, S. Kaplan. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 29, no. 4, Jan. 1952, pp. 223–51.

3. NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN (book), H. W. Bode. D. Van Nostrand Company, Inc., Princeton, N. J., 1945, pp. 185–95.

4. SYNTHESIS OF NON-P.R. DRIVING POINT IMPEDANCE FUNCTIONS USING ANALOG COMPUTER UNITS, W. J. Karplus. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., Sept. 1957.

Self-Balancing Flux-Gate Magnetometers

WILLIAM A. GEYGER
MEMBER AIEE

Synopsis: This paper is concerned with the application of the self-balancing magnetic-amplifier technique in the realm of flux-gate magnetometers. It describes a second-harmonic type of self-balancing flux-gate magnetometer which is operated in conjunction with a push-pull magnetic amplifier in such a way that the d-c flux in the high-permeability cores of the magnetometer is completely balanced by an opposing d-c flux. Thus, the magnetic field to be investigated, e.g., the earth's field in the vicinity of the magnetometer, remains substantially undisturbed. The compensating d-c flux is linearly proportional to the average value of the polarity-reversible unidirectional output current of the magnetic amplifier. With infinite internal gain ("critical regeneration"), this current is varied by an infinitesimal d-c flux increment in the cores of the magnetometer and may be indicated and/or recorded by means of high-precision moving-coil instruments or d-c potentiometer recorders. In a modified arrangement, the self-balancing flux-gate magnetometer is used as a "gradiometer" for measurement of inhomogeneity of magnetic fields.

THE FLUX-GATE magnetometers[1] are those belonging to the large group of magnetically controlled saturable-reactor devices which are based upon the process of magnetizing a high-permeability core by a periodically varying magnetomotive force, and by a unidirectional magnetomotive force, independently and simultaneously. There is obviously a very close linkage between flux-gate magnetometers and magnetic amplifiers because similar magnetic-core arrangements and similar circuit techniques are used. The essential difference between flux-gate magnetometers and magnetic amplifiers consists merely in the fact that in the first case the saturable-reactor device is controlled by the magnetic field to be measured, whereas in the second case this device is controlled by the electric current or voltage to be amplified. Of course, further differences are concerned with the various forms of magnetic-core arrangements, e.g., permalloy cores of flux-gate magnetometers in the form of a rod (open magnetic circuit), and spirally-wound Orthonol tape cores of magnetic amplifiers (closed magnetic circuit).

With regard to basic principle of operation and circuit techniques, flux-gate magnetometers may be classified as follows:

First are the fundamental-frequency type of magnetometers, in which the impedances of saturable-reactor elements in a-c bridge or differential circuits are varied in accordance with the actual magnitude and direction of the magnetic field to be measured. The fundamental principle of these magnetometers is similar to that of magnetic amplifiers utilizing either external feedback or self-saturation.

Second are the second-harmonic-type magnetometers, in which saturable-transformer elements acting as a magnetic frequency doubler produce a second-harmonic voltage which corresponds with regard to magnitude and phase relationship to a second-harmonic reference voltage to the actual magnitude and direction of the magnetic field to be measured.

G. Barth[2,3] introduced various forms of fundamental-frequency-type flux-gate magnetometers utilizing either external feedback or self-saturation. The aim of his research and development work consisted in applying improved magnetic-amplifier techniques in the field of flux-gate magnetometers in order to achieve high sensitivity and the possibility of multistage design with succeeding magnetic-amplifier stages. These full-wave and half-wave flux-gate magnetometers are very similar to the various forms of external-feedback and self-saturating circuits, as used in the magnetic-amplifier art. They may be operated either in the conventional way from an a-c power supply, or in some modified arrangements from a d-c power supply in connection with a mechanical-switch device acting as a chopper or modulator.

Considerable effort has been expended in the study and development of second-harmonic-type flux-gate magnetometers. In 1928, Aschenbrenner and Goubeau[4] applied such a device for recording the changes in intensity of the earth's field. The second-harmonic-type flux-gate magnetometer is an excellent direction finder

---

Paper 58-96, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 11, 1957; made available for printing November 8, 1957.

WILLIAM A. GEYGER is with the U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Md.

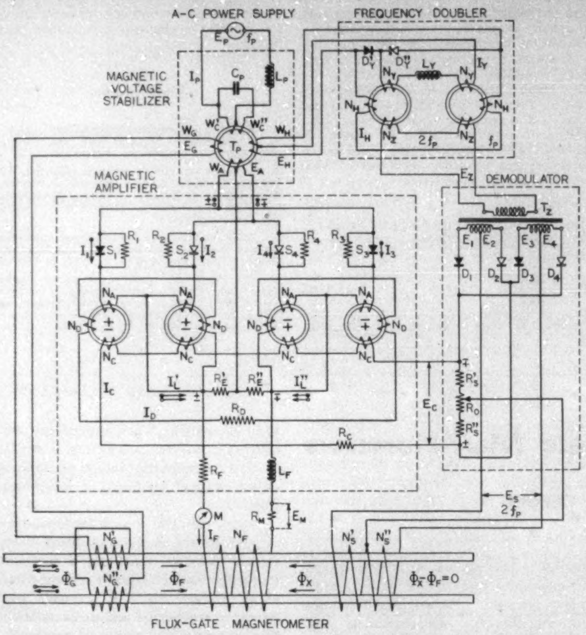


Fig. 1. Circuit diagram of a second-harmonic-type self-balancing flux-gate magnetometer

and may be used as a compass (permalloy compass). This type of flux-gate magnetometer is particularly suitable for geophysical measurements and special military magnetic air-borne detector applications. There are also numerous other possibilities for taking advantage of the very high sensitivity of second-harmonic-type flux-gate magnetometers, e.g., detection of iron particles in aluminum or copper rods, automatic compensation of the earth's field in test rooms, etc. A survey on this subject with references to the literature has been given elsewhere.[1]

This paper is concerned with the possibility of operating a flux-gate magnetometer in such a way that the d-c flux in the high-permeability cores is completely balanced by an opposing d-c flux so that the magnetic field to be investigated, e.g., the earth's field in the vicinity of the magnetometer, remains substantially undisturbed. In a typical example of a 400-cycle-operated self-balancing flux-gate magnetometer, a conventional self-saturating magnetic-amplifier push-pull circuit is d-c controlled from a second-harmonic-type flux-gate magnetometer in such a way that the output current of the magnetic amplifier is varied by an infinitesimal d-c flux increment in the cores of the magnetometer. With infinite internal gain (critical regeneration), the average value of the polarity-reversible unidirectional output current of the magnetic amplifier is linearly proportional to the d-c flux to be measured (the linearity error is of the order of 1/10 of 1%). This output current may be indicated and/or recorded by means of high-precision moving-coil instruments or d-c potentiometer recorders.

In accordance with the specific properties of negative-feedback amplifiers, the accuracy of self-balancing flux-gate magnetometers is substantially independent of changes of the characteristics of circuit components, large variations in magnitude and frequency of the a-c power supply, and load-circuit resistance. Besides these special features, self-balancing flux-gate magnetometers offer the general advantages of magnetic amplifiers and similar devices with transformerlike structures, i.e., elimination of movable parts and vacuum-tube components, sturdiness and reliability, almost unlimited life, etc.

Measurement Principle

Fig. 1 shows the circuit diagram of a second-harmonic-type self-balancing flux-gate magnetometer having two equally rated saturable-transformer elements with high-permeability (e.g., supermalloy) cores, in the form of a rod ("open" magnetic circuit). It is to be noted that the longitudinal dimension of these cores is great in proportion to the transverse dimension. The primary windings N_G', N_G'' of the saturable-transformer elements are supplied from the secondary winding W_G (voltage E_G, frequency f_P) of a power-supply transformer T_P, which also feeds the magnetic amplifier and the frequency doubler, which provides the double-frequency reference voltage E_Z of the demodulator acting as a phase-sensitive rectifier with the operating frequency $2f_P$.

To obtain extreme stability of operation, it may be advisable to employ a magnetic voltage stabilizer (ferroresonant circuit) consisting of a nonlinear reactor T_P with almost rectangular magnetic characteristic, a capacitor C_P, and a linear reactor L_P. In this case, a square-wave voltage E_G (frequency f_P) is applied to the primary windings N_G', N_G'' of the saturable-transformer elements; and square-wave voltages E_A, E_H are applied to the magnetic amplifier and the frequency doubler with demodulator.

The operation of the second-harmonic-type flux-gate magnetometer is based on the differential magnetization of the two saturable-transformer elements by the d-c flux Φ_X to be measured acting in conjunction with the a-c ampere-turns of the series-opposing connected primary windings N_G', N_G'' (a-c flux Φ_G). When $\Phi_X=0$, then the equal and opposite voltages induced in the common center-tapped secondary winding $N_S'N_S''$ are identical in waveshape, and they balance each other in their effect upon the demodulator input circuit ($E_S=0$). However, when a d-c flux Φ_X is applied, the voltages induced in $N_S'N_S''$ become asymmetrical, and a second-harmonic voltage E_S ($2f_P$) appears on the demodulator input circuit. It is to be noted that the phase of E_S (the phase relationship with regard to the second-harmonic reference voltage E_Z) reverses when the direction of Φ_X is reversed.

The graphic symbols of Fig. 1 (black arrow and white arrow indicating the direction of the a-c flux Φ_G in the saturable-transformer elements) show that the d-c flux Φ_X aids the flux component of the first primary winding N_G' and opposes that of the second primary winding N_G'', or vice versa, during the half-cycle periods of supply voltage E_G. As a result, two highly asymmetrical voltage components are induced in the secondary winding $N_S'N_S''$, and a phase-reversible second-harmonic voltage E_S will be applied to the demodulator input circuit. This voltage E_S is, within certain limits, proportional to the actual magnitude of the d-c flux Φ_X. In the arrangement of Fig. 1, a "star-modulator" full-wave type of phase-sensitive rectifier[5] with second-harmonic reference voltage E_Z is employed for converting the phase-reversible second-harmonic voltage E_S across $N_S'N_S''$ into a polarity-reversible d-c signal voltage E_C, which is applied to the control circuit of the magnetic amplifier.

It has been pointed out previously[6] that in a second-harmonic-type flux-gate magnetometer, the d-c flux to be measured may be completely compensated by the opposing d-c flux of a variable direct current, preferably in such a way that the arrangement will be automatically balanced by means of conventional self-balancing instruments, as used in the technique of self-balancing d-c potentiometers. In the arrangement of Fig. 1, the flux-gate magnetometer and a push-pull self-saturating magnetic amplifier with infinite internal gain are operated so that the d-c flux Φ_X to be measured is completely balanced by an opposing d-c flux Φ_F which is linearly proportional to the average value of the polarity-reversible unidirectional output current I_F of the magnetic amplifier. Actually, when employing a properly designed magnetic-amplifier circuit with Orthonol tape cores and silicon junction diodes, the d-c flux $(\Phi_X-\Phi_F)$ is reduced to about 1/1,000 of its original value (Φ_X). Thus, the magnetic field to be investigated, e.g., the earth's field in the vicinity of the magnetometer, remains substantially undisturbed.

The basic principle of the self-balancing flux-gate magnetometer, Fig. 1, consists in supplying a variable polarity-reversible negative-feedback current I_F to compensating winding N_F so that the d-c flux Φ_X to be measured is completely balanced by the opposing d-c flux $\Phi_F=\text{constant}\times I_FN_F$. In this case, $\Phi_X-\Phi_F=0$, and the second-harmonic voltage E_S across secondary winding $N_S'N_S''$ is zero. The relationship between Φ_X and the average value of I_F is linear (the linearity error is of the order of 1/10 of 1%, in accordance with similar performance of self-balancing magnetic amplifiers[7,8]). As indicated in Fig. 1, a choke coil L_F is inserted in series with negative-feedback winding N_F to prevent second-harmonic current from flowing through this winding.

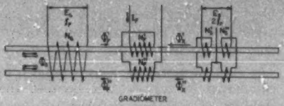


Fig. 2. Arrangement of windings of a self-balancing flux-gate magnetometer for measurement of inhomogeneity of magnetic fields: "gradiometer"

The negative-feedback current, $I_F=\text{constant}\times\Phi_X$, may be indicated and/or recorded by means of high-precision moving-coil instruments or d-c potentiometer recorders. When employing additional magnetic-amplifier stages, the voltage drop $E_M=I_FR_M=\text{constant}\times\Phi_X$ may be applied to the control circuit of a self-balancing magnetic amplifier operating under voltage-balance conditions. Another possibility consists in applying the negative-feedback current I_F to the control windings of a current-type self-balancing magnetic amplifier in a multistage design with succeeding magnetic-amplifier stages.

In accordance with the specific properties of negative-feedback amplifiers, the accuracy of such self-balancing flux-gate magnetometers is substantially independent of changes of the characteristics of circuit components, including the saturable-transformer elements with windings $N_G'N_G''$, N_F, and $N_S'N_S''$. Furthermore, the operation of the flux-gate magnetometer will not be substantially affected by large changes in magnitude and frequency of the a-c power supply (E_P, f_P), and load-circuit resistance (R_F).

In the typical example of Fig. 1, the push-pull self-saturating magnetic amplifier has two equally rated saturable-reactor systems (the plus-minus system and the minus-plus system) with gate windings N_A, d-c control windings N_C, and positive-feedback (boosting) windings N_D providing critical regeneration,[7,8] i.e., infinite internal gain. The resultant voltage drop across series-mixing resistors R_E', R_E'' is proportional to the actual resultant load current $(I_L'-I_L'')$, where $I_L'=I_1+I_2$, and $I_L''=I_3+I_4$. The two saturable-reactor systems are combined in the usual way by means of supply transformer T_P. Quiescent-current values are conveniently adjusted with parallel resistors R_1 to R_4, across silicon junction diodes S_1 to S_4. In accordance with the basic principle of self-balancing magnetic amplifiers, the regeneration-control resistor R_D will be adjusted so that the control-winding current $I_C=E_C/R_C$ is substantially reduced to zero. Proper adjustment of R_D may be achieved easily by means of a microammeter consisting of a lightspot galvanometer with low-resistance shunt (I_C is of the order of ±1 to 2 microamperes): If R_D is adjusted below or above the point of critical regeneration while the cores of the flux-gate magnetometer are subjected to variations of d-c flux Φ_X, a control-winding current I_C will appear but it will be of a direction of current flow dependent upon the direction of actual misadjustment. With critical regeneration, $E_C=0$, $E_S=0$, and $\Phi_X-\Phi_F=0$. In this case, the negative-feedback current I_F indicated by the ammeter M provides full-balance condition[6] of the flux-gate magnetometer: $I_F=\text{constant}\times\Phi_X$.

The star-modulator full-wave type of phase-sensitive rectifier[6] (demodulator) contains the reference-voltage transformer T_Z with two center-tapped secondary windings (voltages E_1 to E_4), four silicon junction diodes D_1 to D_4, a pair of mixing resistors R_S', R_S'', and a potentiometer resistor R_0 for zero adjustment ($E_C=0$ with $E_S=0$). The primary winding of T_Z is supplied from a conventional type of magnetic frequency doubler consisting of two equally rated saturable-transformer elements with Orthonol tape cores, series-opposing-connected a-c windings N_H, series-aiding-connected d-c windings N_Y, and series-aiding-connected secondary windings N_Z, which provide the second-harmonic reference voltage E_Z of the demodulator circuit. The direct current I_Y of the frequency doubler is derived by full-wave rectifiers D_Y', D_Y'' from power-supply transformer T_P; and a choke coil L_Y is used so that the second-harmonic current will flow only through the primary winding of T_Z and will not be short-circuited by the components of the d-c circuit (I_Y).

Modifications

In accordance with common practice, the second-harmonic-type flux-gate magnetometer may be applied with modified arrangement of its windings in such a way that the second-harmonic voltage of the secondary winding is linearly proportional to the actual difference between the d-c flux values of the two high-permeability core elements. This differential-type flux-gate magnetometer, sometimes termed "gradiometer," measures the actual difference between the flux in the first core element and that in the second core element. In this modified form, a self-balancing flux-gate magnetometer may be employed for measurement of inhomogeneity of magnetic fields.

Fig. 2 illustrates the possibility of applying the arrangement of Fig. 1 as a gradiometer, merely by modifying the connections of the winding units of the high-permeability core elements. The flux difference $\Phi_X'-\Phi_X''$ to be measured is compensated by the negative-feedback current I_F flowing through the series-opposing-connected compensating-winding units N_F' and N_F''.

In the example of Fig. 2, it has been assumed that the flux Φ_X' in the first core element with winding N_F' is larger (9 units) than the flux Φ_X'' in the second core element with winding N_F'' (3 units), as illustrated by the different lengths of the arrows representing Φ_X' and Φ_X''. Since N_F' and N_F'' are wound in opposite directions, the negative-feedback current I_F will subtract a compensating flux Φ_F' (-3 units) in the first core element and will add an equal amount of compensating flux Φ_F'' ($+3$ units) in the second core element. Thus, in this example, with full-balance conditions, the first resultant flux $\Phi_X'-\Phi_F'=9-3=6$ units will be equal to the second resultant flux $\Phi_X''-\Phi_F''=3+3=6$ units, and the second-harmonic voltages across the secondary winding units N_S', N_S'' will be zero. Therefore, with critical regeneration of the magnetic amplifier, the negative-feedback current I_F is linearly proportional to the actual flux difference $(\Phi_X'-\Phi_X'')$. If the external field to be investigated is homogeneous $(\Phi_X'=\Phi_X'')$, then the negative-feedback current $I_F=0$. However, when there is a certain amount of inhomogeneity $(\Phi_X' \gtrless \Phi_X'')$, then $I_F=\text{constant}\times(\Phi_X'-\Phi_X'')$.

Other possible modifications are concerned with the design of flux-gate magnetometer cores (e.g., with closed magnetic circuit), magnetic amplifier, frequency doubler, and demodulator. Obviously, many other forms of these devices may be used equally well. A permanent-magnet-type frequency doubler[9] has been used successfully, and other forms of phase-sensitive rectifiers[5] may be applied in such arrangements. The magnetic voltage stabilizer may be eliminated in many cases because the stability of operation is inherently very good as a result of the large amount of negative feedback. It is, of course, also possible to supply self-balancing flux-gate magnetometers from a magnetic frequency multiplier or from a switching transistor d-c to a-c converter[10] to reduce the size of the saturable-transformer and saturable-reactor elements, and to increase the speed of response. Finally, it is to be noted that the basic principle of self-balancing flux-gate magnetometers may be applied equally well to those fundamental-frequency-type flux-gate magnetometer circuits[1–8] which are very similar to the various forms of external-feedback and self-saturating circuits, as used in the magnetic-amplifier art.

Conclusions

1. The combination of a second-harmonic-type flux-gate magnetometer and a push-pull magnetic amplifier with infinite internal gain makes it possible to balance the d-c flux in the high-permeability cores of the magnetometer by an opposing d-c flux which is linearly proportional to the average value of the polarity-reversible unidirectional output current of the amplifier. In this case, with critical regeneration, the self-balancing flux-gate magnetometer is operated in such a way that the duodirectional output current of the magnetic amplifier is varied by an infinitesimal d-c flux increment in the core elements of the magnetometer.

2. When employing a properly designed magnetic-amplifier push-pull circuit with Orthonol tape cores and silicon junction diodes, the d-c flux in the cores of the flux-gate magnetometer is reduced to about 1/1,000 of its original value. Thus, the magnetic field to be investigated, e.g., the earth's field in the vicinity of the magnetometer, remains substantially undisturbed.

3. The relationship between the d-c field to be measured and the average value of the polarity-reversible unidirectional output current of the magnetic amplifier is linear (the linearity error is of the order of 1/10 of 1%, in accordance with similar performance of self-balancing magnetic amplifiers). This output current may be indicated and/or recorded by high-precision moving-coil instruments or d-c potentiometer recorders.

4. In accordance with the specific properties of negative-feedback amplifiers, the accuracy of measurements with self-balancing flux-gate magnetometers is substantially independent of changes of the characteristics of circuit components including the saturable-transformer elements of the flux gate. Furthermore, the operation will be substantially unaffected by large changes in magnitude and frequency of the a-c power supply, and by load circuit resistance.

5. Besides these special features, self-balancing flux-gate magnetometers offer the general advantages of magnetic amplifiers and similar devices with transformer-like structures, i.e., the elimination of movable parts and vacuum-tube components, sturdiness and reliability, almost unlimited life, etc.

References

1. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957.

2. Method and a Device for Determining the Magnitudes of Magnetic Fields, G. Barth. *U.S. Patent no. 2,252,059*, Aug. 12, 1941.

3. Means for Measuring Magnetic Fields, G. Barth. *U.S. Patent no. 2,390,051*, Dec. 4, 1945.

4. An Arrangement for Recording of Fast Magnetic Disturbances (in German), H. Aschenbrenner, G. Goubeau. *Hochfrequenztechnik und Elektroakustik* (Jahrbuch der drahtlosen Telegraphie und Telephonie), Leipzig, Germany, vol. 47, no. 6, 1936, pp. 177–81.

5. Geyger, *op. cit.*, pp. 326–30.

6. *Ibid.*, pp. 272–73.

7. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 383–95.

8. Self-Balancing Magnetic Amplifier, W. A. Geyger. *U.S. Patent no. 2,700,130*, Jan. 18, 1955.

9. Geyger, *op. cit.*, pp. 274–75.

10. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

An Improved 6-Channel Time Division Multiplex for Submarine Cable Telegraphy

H. F. WILDER
MEMBER AIEE

THE DEVELOPMENTS of recent decades have sustained the prediction made in 1890 by Mahan that if a successful transisthmian canal were to be built in Central America, the Caribbean would become a great commercial area and a crossroads of transportation, not only of local and of South American traffic but between Europe and Asia as well.[1] The importance of Havana, Cuba, in the day of the sailing ship, seated as it is in a geographical position so favored by winds and ocean currents as to command the only passage to Europe, has not been diminished but enhanced by the prophesized burgeoning trade and intercourse. To meet the communications needs of this market place and the richly cultivated area to which it gives access, the Western Union Telegraph Company operates three submarine telegraph cables between Havana and Key West, Fla., an average distance of 101 nautical miles. The second cable of the three was laid in 1889, almost as if in anticipation of Mahan's observation, and the third was laid in 1917, three years after the completion of the Panama Canal and, seemingly, in confirmation of it. Until recently, these duplex-operated cables were terminated in differentially connected polar relays reflecting the land-line practice of those decades, a condition sustaining 3-channel multiplex operation or a signalling capacity three times that of the earlier Morse circuits.

Lines to Cuba

Probably no other division of the submarine cable network of the company presents as many diversified traffic requirements as these connections to Cuba and their extensions to South America and the islands of the Caribbean. One requirement, automatic exchange of traffic with the recorder code-operated cable of Cable and Wireless, Ltd., to San Juan, Puerto Rico, was effected by translating the 3-position signals, dot, dash, or space, to 2-element signals of twice the line frequency, and superposing the latter on appropriate segments of one of the New York 3-channel multiplexes. Now, a second interconnection of this nature is in operation. This circuit, the Island Chain, finally terminates in Barbados, B.W.I.

Another early requirement supplied by the company was the transmission to brokerage houses in Havana of the quotations of the New York Stock Exchange. This service was extended from the mainland in 1930 by superposing on one of the three physical circuits a unidirectional single-channel amplitude-modulated carrier system.

Here then, was further indication of the tide of specialized traffic demands arising from the operations of mercantile houses and the international airlines; the mercantile houses trading in sugar (the staple product of the island), coffee, tobacco, fruit, and chocolate. To provide these leased facilities a second carrier system was designed and put into operation in 1949.[2] Each of the two additional frequency-modulated duplex channels so derived is capable of carrying either a teleprinter circuit or a triple-channel multiplex. And in 1950, the older multiplexes were replaced by similar equipment which was capable, however, of accepting and delivering teleprinter start-stop code signals from and to the customer extensions.

Signal-Shaping Amplifiers

The somewhat higher signalling speed of the new multiplexes highlighted the need for a cable receiving device of greater sensitivity and stability than the unaided polar differential relay. Consequently, in 1955 all the Havana cables, together with the Tricore cable between Key West and Miami Beach, were equipped with signal-shaping vacuum-tube amplifiers.

These amplifiers are in widespread use throughout the submarine cable network of the company, to equalize the transmission distortion and to amplify the signals on cables from 50 to 500 miles in length. The output relays are driven by a vacuum-tube amplifier and while earth current potentials and corresponding low-frequency components in the residual duplex unbalance are prohibited from influencing the output because of the discrimination inherent in the equalizing networks, low-frequency signal components equivalent to the suppressed disturbing currents are restored to the signal by a self-energized correction circuit.

High-Speed Signalling

At the time the amplifiers were installed, tests were made to determine the feasibility of higher signalling speeds, for again the pressure for additional channels to Havana was being felt and it was thought that the new amplifiers might be the key to a full exploitation of the latent signalling spectrum of these low-attenuation cables. For the investigation of higher speeds, magnetic tapes were prepared in the laboratory. These contained the signal combinations that would produce the same transmission distortion and attenuation to be encountered at the speed of signalling of a 6-channel multiplex, at 100 cycles per second, compared to the usual 50 cycles, and these tapes were then inserted in a magnetic tape-recorded signal transmitter at Key West for observation at Havana.[3] A signal of good quality was obtained at the higher speed by readjustment of the signal-shaping network in the terminal amplifier, and the duplex balance was easily improved to a workable accuracy. Despite the greater amplification, disturbance from natural sources and transmission on other cables was still negligible, and when the frequency-modulated carrier was superposed on the test cable without mutual interference, practical operation of a 6-channel duplex multiplex circuit on any of the three cables seemed assured.

The success attending these tests represented a real breakthrough in the previous restriction on channels to Havana. The higher signal capacity of 6-channel multiplexes and their compatibility with

---

Paper 58-44, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 11, 1957; made available for printing December 3, 1957.

H. F. Wilder is with the Western Union Telegraph Company, New York, N. Y.

The author wishes to recognize the contributions of the several engineers whose special efforts made the 6-channel idea realizable. S. A. Kirkowski designed the new start-stop distributors, and W. J. Ramhorst, as a continuation of previous submarine cable equipment design, furnished the physical layouts of the racks. Synthesis of the basic circuitry and the carrying of the project to an operative conclusion was done by G. T. Fontaine.

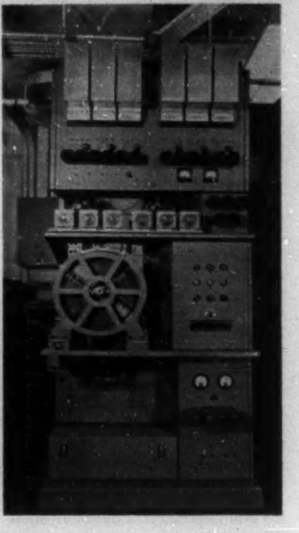


Fig. 1. The sending 6-channel multiplex terminal

the frequency-modulated carrier meant that the number of multiplex channels could be doubled simply by the provision of new terminal equipment, a step comparable to the doubling of transatlantic facilities that had been effected by the insertion of submerged repeaters.

Multiplex Design

Design of the 6-channel multiplex terminals to use the wider signal spectrum now known to be available was soon initiated. Aside from designing for operational stability and the use of well-tried features familiar to the operating personnel, several fundamental decisions were made. In view of the large number of channels ultimately possible, it was thought desirable to avoid use of tape reperforation to couple incoming teleprinter signals to the cable time-division channels. The untidy tape loops and the maintenance of tape continuity were eliminated by the incorporation of a channel repeater between each teleprinter signal-receiving start-stop distributor. Correlated with the decision to use channel repeaters was the limitation of the teleprinter extensions to a speed of 368 operations per minute, nominally referred to as 60-word operation. To have designed for 65 words or 390 operations per minute would have meant the transmission into the domestic telegraph system of teleprinter signals containing rest or stop pulses only 86% of unit length, and this would have precluded regeneration at repeater points, thus rendering the longer teleprinter signal extensions much more susceptible to distortion and bias.

The Sending Terminal

The sending terminal is shown in Fig. 1 and the flow of intelligence through this equipment may be seen in Fig. 2. Each sending rack provides means for coupling six incoming teleprinter lines to their respective 5-unit code multiplex channels. The individual relay banks housing the channel repeaters are mounted at the top of the rack with the teleprinter circuit line and selector relays beneath them. A new start-stop distributor was designed to receive the incoming teleprinter signals from each extension and to deliver the five intelligence impulses to one of two capacitor-storing banks in the associated channel repeater.

A group of three units is driven by a common motor geared to produce the correct brush speed, the engagement of the brush shaft being effected by a dry metallic clutch of recent design significantly free of variation in take-up lag. The brushes are shown at rest on the stop segments of the faceplate; see Fig. 3. Upon the reception of a teleprinter character the clutch magnet circuit is energized by the start impulse of spacing polarity preceding the five intelligence impulses. The brush then rotates for a single revolution, returning to the rest segment coincident with the arrival of the stop impulse of marking polarity at the end of the character. Here it remains until the start impulse of the next character occurs. During its revolution the brush passes over five short-length scanning or pickup segments at intervals of 22 milliseconds. The successive impulse polarities of the five intelligence impulses thus selected are momentarily stored on a relay. The longer segment following each selection is used to deliver a charge from the armature of the relay to an associated capacitor in one or the other of the two capacitor-storing banks in the channel repeater. The stored charges are subsequently commutated by segments on the multiplex faceplate to the coils of five storing relays. After positioning, the armatures of these relays are given time to settle firmly on their contacts before delivery of the signals to the sending amplifier.

To obtain the sixth pulse segment for operation of the transfer relay in the

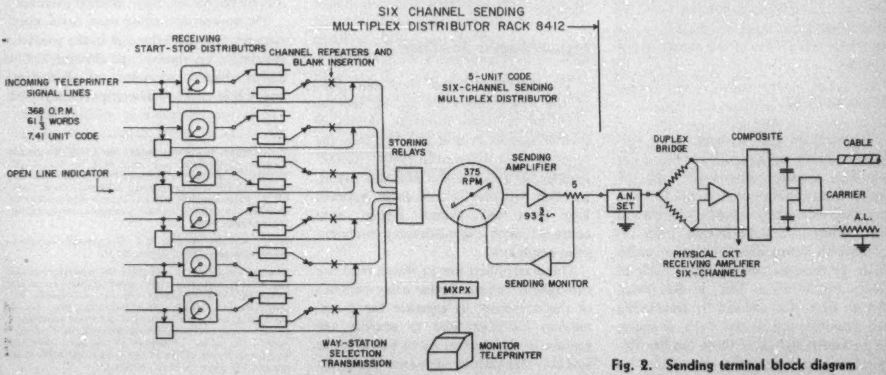

SIX CHANNEL SENDING
MULTIPLEX DISTRIBUTOR RACK 8412
RECEIVING
START-STOP DISTRIBUTORS
CHANNEL REPEATERS AND
BLANK INSERTION
INCOMING TELEPRINTER
SIGNAL LINES
368 O.P.M.
61 1/3 WORDS
7.41 UNIT CODE
OPEN LINE INDICATOR
5-UNIT CODE
SIX-CHANNEL SENDING
MULTIPLEX DISTRIBUTOR
STORING
RELAYS
375
RPM
SENDING
AMPLIFIER
5
93 3/4 ~
A.N.
SET
DUPLEX
BRIDGE
COMPOSITE
CABLE
CARRIER
A.L.
PHYSICAL CKT
RECEIVING AMPLIFIER
SIX-CHANNELS
SENDING MONITOR
MXPX
MONITOR
TELEPRINTER
WAY-STATION
SELECTION
TRANSMISSION

Fig. 2. Sending terminal block diagram

Fig. 3. Faceplate of receiving start-stop distributor

channel repeater, it was necessary to shorten the fifth delivery segment. Closure of the sixth pulse segment shifts the delivery of the next character received to the alternate bank. Both the sixth and stop segments are isolated by short segments. By means of an indicator knob, the faceplate may be shifted to measure the quality of the incoming teleprinter signal and to center the selection point midway between the minimum limits of variation in signal impulse arrival time.

The large multiplex distributor shown directly below the shelf of six start-stop distributors is already in widespread use as a time division commutator on the North Atlantic system. For use on the Key West to Havana cables, the distributor was equipped with a faceplate possessing on one commutator all the segmented rings necessary to both sending and receiving functions, an interchangeability that reduced the number of units to be manufactured and held in reserve at each station. One ring is segmented to provide 22-millisecond timing intervals for each one of the six channels while either translating the multiplex signals to teleprinter signals at a receiving terminal, or monitoring the signal transmitted to the apex of the cable duplex bridge at the sending terminal. The other rings are segmented in the conventional manner for selection, correcting, transmission of unit length impulses, and local functions.

Channel Repeating

The distributor brushes must rotate slightly faster than the maximum rate of transmission of the incoming teleprinter signals so that, regardless of the normally encountered power frequency changes in the supply to teleprinter keyboard and transmitter-distributor motors, the distributor will never fall behind in phase, but will from time to time "catch up" with the signal input. To stabilize the speed at 375 rpm or at 7 rpm above the maximum teleprinter character rate, the usual sliding weights were removed from the distributor driving fork and the fork inclosed in a draft and sunlightproof cabinet located on the plinth of the rack. With this slight speed differential, the signal attenuation is not appreciably increased and the rest or stop pulse duration of each translated teleprinter character issuing from the receiving rack is still well in excess of unit length. At 375 rpm the length of this pulse is 28 milliseconds or 27% greater than a unit-length impulse of 22 milliseconds. It might be said that the code retransmitted is one of 7.27 units instead of the 7.41-unit code signals originally incoming to the sending rack. Regeneration of the 7.27-unit code signal characters may be effected at one or more repeater points without "clipping" the rest pulse.

About every 8 or 9 seconds, the slight amount of excess multiplex speed causes the brush-extracting intelligence stored in one bank of capacitors to approach the segments associated with the alternative bank. Before a scramble can occur, however, an auxiliary or auto-stop circuit continuously searching for this condition, stops the alternation of the transfer relay feeding information to the multiplex for a single character and substitutes a blank for transmission. Thus a new lag, equivalent to the time duration of a character, is introduced and the cycle repeats. If the power frequency at the sending extension increases, the rate of blank insertion decreases; it increases if the frequency drops below 60 cycles per second. Should the teleprinter extension send sporadically as from a keyboard, more than one blank may be transmitted between characters and in the limit where sending ceases the blank is transmitted continuously.

Way-Station Selection

Leased circuits frequently contain one or more drops or way stations in addition to the terminals and are then provided with way-station selecting apparatus enabling one station to call any other. Station selections are preceded by an open-line condition lasting a few seconds, a signal serving to start all teleprinter motors in anticipation of the transmission of the selection character group itself. Ordinarily, a continuous spacing signal would be sent over a channel of a 5-unit code multiplex as a series of "blanks" equivalent to the idling condition, and the significance of the preselection open-line interval would be lost. The fact that the upper-case shift function or "letters" character is sent but once in traffic operations presented an opportunity to translate the open line interval to a form acceptable to the multiplex. Accordingly, whenever a multiplex channel is assigned to the support of a leased circuit operative with way-station selection, the usual open-line alarm circuit is arranged

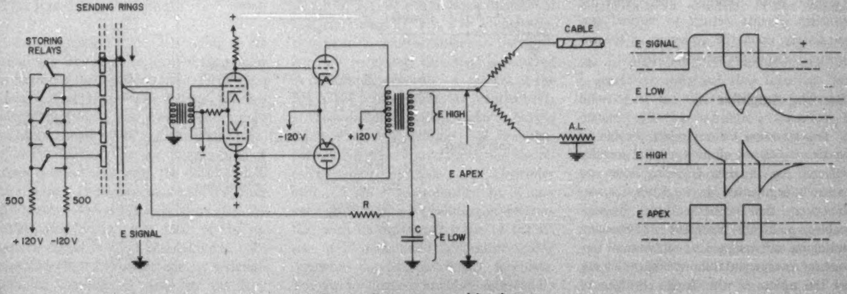


Fig. 4. Transmitting amplifier theory

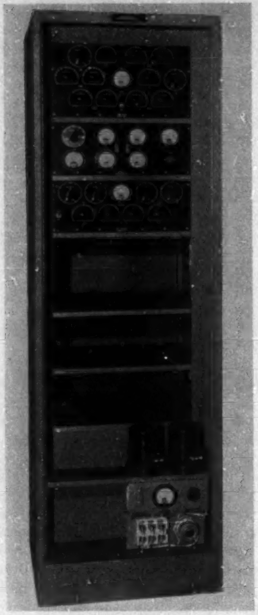


Fig. 5. The receiving signal-shaping amplifier

to send a succession of LTRS (letters) functions (all five impulses of marking polarity) for the 3- to 5-second preselection interval.

At the receiving terminal, the reception of a succession of shift functions, quite in excess of traffic need, is recognized as the preselection open, and in response the terminal apparatus opens the out-going customer extension to activate all teleprinter motors for the call soon to be received. Both sending and receiving sensing circuits return to normal immediately upon the reclosing of the incoming teleprinter circuit extension.[4]

Sending Amplifier

Transmission to a submarine cable at speeds in excess of 50 cycles is usually effected by brushes traveling over accurately segmented rings. It was known, however, that on the three Havana cables, a highly inductive multisection antinoise set designed to suppress all frequency components above 300 cycles lay in the apices of the duplex bridges to prevent physical transmission from interfering with operation of the carrier systems. Since previous experience had demonstrated the impossibility of transmitting from segments to an inductive sending end impedance because of segment gap flashover, it was decided to transmit to the duplex apex from a vacuum-tube power amplifier, which is mounted at the base of the rack to the right of the switch compartment and the sending fork box. To obtain a satisfactory signal shape with the relatively simple networks available in the distant receiving terminal amplifier, however, all frequency components, including the d-c component, must be transmitted to the cable by the sending apparatus. Accordingly, the amplifier was designed to impress only the higher frequency components of the signal upon the duplex apex; the d-c component and the very-low-frequency components are still supplied from the sending segments through a resistive-capacitive network.

Fig. 4 shows the circuit theory of the sending amplifier. Segments of the transmitting ring are protected by unusually large battery tap resistors of 500 ohms so that "bridging" by a worn brush is of no importance. The resistor R of about 1,000 ohms in the LOW path limits the d-c cable current to 40 or 50 milliamperes and the 40-microfarad capacitor C suppresses all high-frequency components while simultaneously providing a low-impedance path to ground for the high-frequency components generated by the vacuum-tube amplifier in the HIGH path. The latter is a push-pull power amplifier transformer coupled to the cable duplex bridge and is driven by a flip-flop vacuum-tube relay that "square up" the otherwise

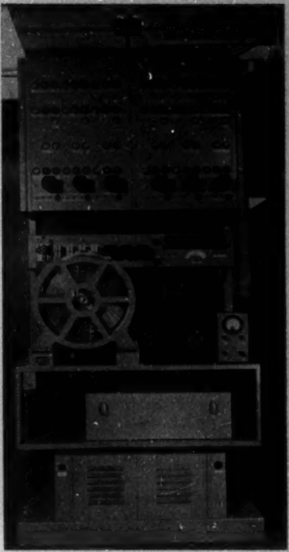


Fig. 6. The receiving 6-channel multiplex terminal

rounded off waveform from the ring caused by the high battery tap resistors. An inspection of the voltage loci shows that the sum of the high- and low-path voltages reproduces at the cable bridge apex a waveform which frequency and

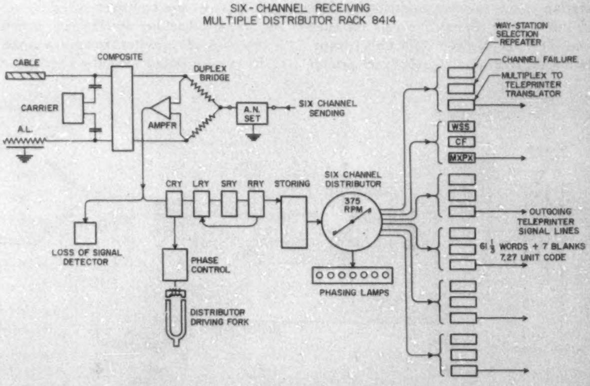

SIX-CHANNEL RECEIVING
MULTIPLE DISTRIBUTOR RACK 8414
CABLE
COMPOSITE
DUPLEX BRIDGE
CARRIER
A.L.
AMPFR
A.N. SET
SIX CHANNEL SENDING
WAY-STATION SELECTION REPEATER
CHANNEL FAILURE
MULTIPLEX TO TELEPRINTER TRANSLATOR
WSS
CF
MXPX
CRY
LRY
SRY
RRY
STORING
SIX CHANNEL DISTRIBUTOR
375 RPM
LOSS OF SIGNAL DETECTOR
PHASE CONTROL
DISTRIBUTOR DRIVING FORK
PHASING LAMPS
OUTGOING TELEPRINTER SIGNAL LINES
61 1/3 WORDS + 7 BLANKS
7.27 UNIT CODE

Fig. 7. Receiving terminal block diagram

delay distortion have been eliminated. The entire system is energized from the station 240-volt d-c power and no special cable rectifiers or batteries are required.

Sending Monitor

Experience has shown that it is imperative that the technician be provided with means to monitor the transmitted signals and that the only reliable method is to drive the monitoring device by the current actually entering the apex of the duplex bridge, for then an interruption of the cable or artificial line currents will announce failure in circuit continuity. At the sending terminal, a 5-ohm resistor is placed in series with the output of the sending amplifier and the voltage drop produced across this resistor by the charging currents to the cable apex after amplification drives a high-speed monitor relay. The multichannel multiplex signal output of the relay is then returned to a vacuum-tube translator which, in conjunction with the teleprinter timing ring on the distributor faceplate, restores the start and stop pulses of a given teleprinter channel to permit the technician to observe the transmission of any one multiplex channel on a 60-word teleprinter. The monitor cabinet and channel selector switch are contained in a cabinet adjacent to the multiplex distributor.

At the request of the distant station received either over the international talk-channel or on the Morse facilities provided on these racks, the sending terminal technician can transmit the "all-idle channel" pattern for phasing of the receiving distributor, or a recurrent pattern which is devoid of characteristic distortion, thus enabling the receiving station technician to range and center the signal pickup without requiring the transmission of a test tape.

The Receiving Terminal

In propagation from the sending to the receiving terminal, the form of the signals is degraded by the effects of frequency and delay distortion. On cables of this electrical length, the former effect may reduced the amplitude of the highest frequency reversals to one twenty-fifth of their original size while the delay distortion retards the time of arrival of the lower frequency components relative to the attenuated highs. The form of the received signal is quite unintelligible. To equalize these distortion effects, a signal-shaping amplifier is used on the Key West–Havana cable and on other cables of the same order of attenuation. Two such amplifiers are usually mounted in a single cabinet although but one is shown in Fig. 5. For speeds of operation of over 50 cycles, high-speed relays whose upper limit of operation is several hundred cycles per second are substituted for those shown in the figure.

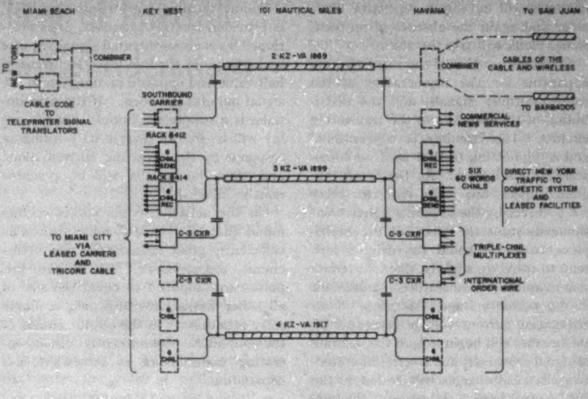


Fig. 8. Ultimate facilities

Adjustable filter sections located in the upper compartment shape the received signals and suppress the very-low-frequency earth current and residual unbalance potentials, and the electrical noise frequencies lying outside the upper end of the signal spectrum. The shaped signal waveform, minus the zero-frequency component, is then amplified by a 2-stage transformer-coupled amplifier driving one or two output relays. The output relay drives the receiving relay loop into the multiplex set and a low-pass filter or local correction network, which restores the d-c signal component.

The 6-channel receiving terminal is shown in Fig. 6, and a block diagram of this equipment in Fig. 7. A single receiving rack complementary to the sending rack previously described is required at each cable terminal. A cabinet at the top of the rack contains six identical sets of vacuum-tube circuitry, each set having three distinct functions. Of first importance is the circuit translating the received multiplex signals to the teleprinter form. This is accomplished by a new vacuum-tube circuit operating in conjunction with its associated group of the six sets of 22-millisecond teleprinter timing segments on the distributor faceplate. The five intelligence impulses received by the multiplex are a little over 5 milliseconds in duration. These are first stored on five capacitors. Each capacitor is then discharged by its respective segment on the teleprinter timing ring through one winding of a transformer having seven primary windings. Of the seven timing segments for each channel, the first segment discharges a positively charged capacitor through the no. 1 winding of the transformer to prefix the start impulse to the teleprinter character being developed. Similarly, a seventh but negatively charged capacitor is provided to produce the stop impulse.

The voltage transients induced in the transformer secondary drive a vacuum-tube bistable amplifier, which in turn positions a relay controlling the outgoing teleprinter line. Auxiliary circuitry holds the line closed whenever traffic ceases.

The way-station selection repeater has been previously described. When not in use, it is switched to sense an open teleprinter line and to bring such an abnormal condition to the attention of the technician.

A channel-failure alarm is the third feature provided. This circuit senses, for a greater-than-normal time interval, the absence of the time differential blanks, caused perhaps by a failure in the channel repeater circuitry or by loss in phase of the receiving distributor. The basis of the circuit is a capacitor charged through a high resistance. If during traffic transmission the time differential blanks do not appear at the usual rate of seven or so per minute the voltage on the capacitor rises excessively and triggers an alarm. Should the distributor drop behind in phase by one or more segments, all idle channels will develop a repetitive character and again the absence of an occasional blank will trip the alarm.

The lower portion of the rack supports apparatus for the regeneration of the cable multiplex signals, and the distribution of the impulses to the translating circuits. The circuitry is conventional and a duplication of that used on transatlantic multiplexes. As the block diagram implies, one relay, corrector relay CRY, driven by the incoming signals, continuously scans the phase of the multiplex brushes. Should the driving fork tend to speed up slightly, due to a reduction in ambient temperature or a decrease in d-c power voltage, the phase of an alternating current voltage generated by the brushes will begin to lead the instants of signal crossover, and a series of corrective electrical charges will be fed to the fork control bank. As a result, the current flowing in a damping coil at the extremities of the fork tines will be decreased thereby lessening the stiffness of the fork and causing it to slow the speed of brush rotation. Conversely the magnetic flux is increased by the same corrective circuits should the fork slow down. Thus, the receiving brushes rotate continuously at precisely the same speed as the sending brushes and in channel phase with them, although because of the time of propagation over the cable, there will be an angular lag equivalent to a fraction of a revolution.

A line relay LRY and a selector relay SRY, operating in conjunction with 30 scanning segments on the faceplate, regenerate the signal impulses and deliver them to 3 storing relays. Storage of the regenerated impulses for a short time permits the scanning segments to be mounted on an orientable ring, enabling the technician to center the range.

On the long transatlantic cables where the amplitude of the most rapid signal reversals is greatly attenuated, the signal-shaping networks of the receiving amplifier are tuned to emphasize the next lowest reversal frequency whose impulses contain two transmitted impulses rather than one. These "doubles" reversals are very much larger than the "singles" and are not degraded by electrical noise. Although under conditions of suppressed singles operation any singly transmitted impulse does not appear in the amplifier output the missing impulse may be interpolated by the use of appropriate circuitry within the multiplex equipment. On the 6-channel terminal a reversing relay RRY is provided for this purpose. It is positioned by each regenerated signal impulse and returns to the line relay a current of half value and opposite in direction to the signal impulse last seen. If the next impulse is a suppressed impulse the line relay will be free to reverse its position in response to the reversing current alone and thus the impulse will be synchronously "filled in."[3]

On the actual Havana cables it was found that the signal-to-noise ratio was sufficiently great to make the use of this circuit unnecessary. The singles impulses are amplified to equal the size of all other frequencies and only a single relay is necessary in the output circuit of the amplifier. Consequently circuit-operating maintenance is somewhat less demanding.

Regulating

When preparing a multiplex facility for traffic use, the attendance of both the sending and receiving terminal technicians is required. First, the sending terminal must be made aware of the need by the receiving terminal for the transmission of the "all channel idle" or "spacing" condition in order that the receiving brushes may be wheeled around until, by observation of the neon phasing lamp bank, channel phase is established. Second, the pattern combination, and a range excursion not less than 50% of the theoretical maximum obtained under duplex conditions must be requested before traffic is accepted. It is necessary therefore that the receiver be able to call in the sending technician, and so each receiving rack is equipped with a loss-of-signal detector. When the need arises, the receiving station simply stops transmission in the opposite direction, thereby immediately alerting the distant sending technican.

Successful Operation

The 6-channel equipment was placed in service early in April 1957. Its immediately successful operation not only provided three additional duplex teleprinter channels to Havana but permitted the 3-channel multiplex then in the old Miami office to be abandoned just prior to the establishment in the new office of more up-to-date operations through Atlanta. A duplicate set of 6-channel equipment is now under construction and other plans are afoot to increase the channels potentially available from the three cables. In Fig. 8, the author indulging in a certain amount of latitude, has sketched the application of the new 6-channel equipments, and the projected carrier channel terminals and recorder code combining apparatus to this communication network. As the need arises, the ultimate realization of a total of some 27-odd duplex and commercial news circuits to Cuba is thus seen in the offing.

A second frequency-modulated carrier shown superposed on cable 4 will be a duplicate of the original; the performance of the latter could scarcely be improved upon, while the replacement of the "old work-horse," the amplitude-modulated carrier, by a 3-channel unidirectional carrier transmitting commerical news service will relieve the existing triple multiplex of the inefficient use of their otherwise duplex operative channels. Similarly, the completion of the design and the installation of special time division apparatus, to combine over a single physical or carrier channel the two duplex recorder code circuits interconnecting with the Cable and Wireless, Ltd., system in the West Indies, will further increase the efficiency of use of the existing 3-channel multiplexes to New York. The bidirectional recorder code to teleprinter signal translators associated with this mode of cable code operation are in the process of development. This apparatus will not only translate the alphanumerical characters of one code to the electrical impulses of another code but will substitute for the special designations which are almost traditional in recorder code the equivalent function characters common to reperforation switching systems. And southbound, other special apparatus adjunct to the translators will automatically inject the correct "attack" or way-station call of the many stations along the island chains.

References

1. THE INFLUENCE OF SEA POWER UPON HISTORY (book), **A. T. Mahan.** Pp. 33, 315.

2. A 2-CHANNEL CARRIER TELEGRAPH SYSTEM FOR SHORT SUBMARINE CABLES, **E. L. Newell, C. H. Cramer.** *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 61–65; also *Western Union Technical Review*, New York, N. Y., vol. 4, no. 2, Apr. 1950.

3. A MAGNETIC TAPE SIGNAL TRANSMITTER FOR OCEAN CALBES, **E. T. Finnegan.** *Western Union Technical Review*, vol. 10, no. 3, July 1956.

4. A COMBINATION MULTIPLEX-TELEPRINTER WAY-STATION CIRCUIT, **H. F. Krantz II.** *Ibid.*, vol. 11, no. 3, July 1957.

Vol.
Pt. 1
1958

Nation-wide Switching of Intertoll Trunks—Transmission Considerations

O. F. WALLMAN
MEMBER AIEE

THE successful planning, installation, and operation of a nation-wide switching plan to interconnect more than 62 million telephones in the United States requires a high degree of teamwork between various groups of people. This involves the development, design, transmission, equipment, and building engineers to provide the system and the operating and maintenance people to keep it going. This paper deals primarily with transmission phases of switching telephone trunks in the United States. Brief mention is made of the interconnection with the Canadian switching plan involving approximately 5 million additional telephones.

A 1953 AIEE paper covered the fundamentals underlying the design of intertoll trunks.[1] This paper makes use of the basic data but covers primarily transmission considerations involved in switching of intertoll trunks in the nation-wide switching plan.

A new nation-wide switching plan employing switching machines instead of manual interconnection of intertoll trunks by operators was introduced in the Bell System in 1950.[1-5] The original plan was based on the use of 4-wire switching at all major switching offices and 2-wire switching at toll centers, the junction points between intertoll and exchange facilities.

Since its introduction a number of changes in concept have already been introduced into the plan as follows:

1. Discontinuance of the national center.
2. Introduction of 2-wire switching systems at many of the smaller switching offices.
3. Integration of the switching plans of the United States and Canada in a continent-wide switching plan.
4. New techniques in the assignment and operation of echo suppressors to meet the needs of the expanded switching plan.

A measure of the magnitude and complexity of the transmission problems involved can best be seen from the number of switching points involved. The present view indicates an ultimate total of about 300 major switching offices and about 2,000 toll centers. The plan must be flexible enough to permit interconnection of the facilities in the United States with the European systems through the newly installed transatlantic submarine telephone cable and with the Hawaiian Telephone System through the transpacific submarine telephone cable.

The basic plan for switching together the necessary number of intertoll trunks to permit the interconnection of any two telephone customers in the United States is shown in Fig. 1. It will be noted that the maximum number of trunk circuits in any one connection is nine for a telephone call within the United States. As outlined later, some calls between Canada and the United States may contain a total of ten links. Two of these links, one at each end, connect the customer's telephone from his local central office to the toll network at the toll center. Two such links are in every connection except for those relatively small number of cases where direct intertoll trunks are provided between two local offices in different cities. Their design with relation to return losses in the frequency ranges important from an echo and singing standpoint is covered in another paper.[6] It will be noted in Fig. 1 that the switching offices are given both names and numbers to indicate their rank or class in the switching plan. Some people prefer the names while others prefer numbers in referring to the various trunk links. Each switching office below the rank of "regional center" is connected to one of higher rank by what is known as a final trunk group (shown by the solid lines). Each regional center is connected by a final group to each other regional center in the United States. In addition, high-usage trunk circuits (shown in Fig. 1 by dashed lines) are employed between the various switching systems as determined by economic studies. With switching machines capable of automatically selecting as their first choice a direct trunk, if available, and of making successive choices involving an increasing number of links, the number of links in a connection may vary between one and eight for any two customers, in the same regional-center area. For calls between customers in different regional-center areas the number of links may vary between three and nine. The maximum use of high-usage trunks, as permitted by economic studies, results in a reduction in the number of links in a telephone connection and thereby in improved transmission performance.

Fig. 2 shows the present view of the location of the ten regional centers (RC) in the United States and of the two in Canada. This indicates one addition and the relocation of three others as compared to the previous view. The number and location of regional centers are based on studies to obtain an approximate cost balance between intertoll trunk and switching costs. These studies are a continuing project and changes and additions in the number and location of regional centers or other switching centers will be made in years to come. The transmission plan must, therefore, be reviewed continuously and be flexible enough to follow the indicated economic changes. The Canadian and United States switching plans are integrated by means of intertoll trunks between the two Canadian regional centers and some of the regional centers in the United States. Full interconnection on a final basis is at present uneconomical so that for some calls between Canada and the United States there will be ten links involved in the connections. This results in RC-RC-RC connections for those cases where final groups have not been provided.

Basic Transmission Considerations

Before discussing the various switching systems employed it will be desirable to review briefly the following most important transmission considerations involved in nation-wide switching of intertoll trunks.

1. Crosstalk.
2. Noise.
3. Distortion (frequency bandwidth).
4. Echo and echo control.
5. Singing or near singing.
6. Intertoll trunk losses (deviations from assigned losses).

The first two factors, crosstalk and noise, are controlled by the design of the intertoll trunks. This involves primarily the spacing of carrier and voice-frequency repeaters. It is important that the fundamental design of the system meets the transmission requirements for its

---

Paper 58-5, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 7, 1957; made available for printing October 11, 1957.

O. F. Wallman is with the American Telephone and Telegraph Company, New York, N. Y.

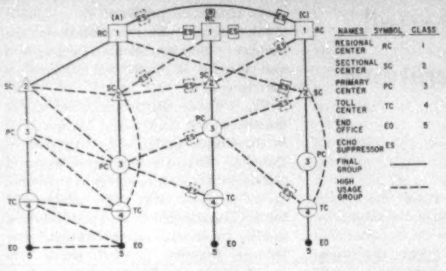


Fig. 1. Interconnection of high-usage and final intertoll trunk groups

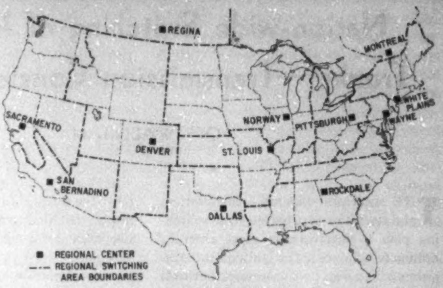


Fig. 2. Regional area boundaries: tentative April 1957 view of 1965 arrangements

intended use in the nation-wide layout. The design of the switching system components, including wiring and cabling, must be such as to provide negligible contribution to over-all noise and crosstalk.

Amplitude (or transmission) distortion is also a fundamental consideration in the design of a carrier or voice-frequency system. This involves the provision of sufficient bandwidth of the channel filters and associated terminal equipment to realize the full capabilities of the telephone instruments. An important factor in this respect is to hold the amount of office-type cable (series or multiple) to reasonably short lengths. The amount contributed by any one switching office may be relatively small but the accumulative effect of the wiring in each of the switching offices shown in the built-up connections in Fig. 3 may result in considerable transmission distortion in long built-up connections unless careful attention is given to the cabling lengths in the individual offices. Office-type cabling generally has a high mutual capacitance (about 0.125 microfarad per mile). Tests indicate that an office (900 ohms impedance) whose cabling (series and multiple) measures 0.07 microfarad will have a loss 1 db (decibel) higher at 3,000 cycles than at 1,000 cycles on a switched connection.

Echo and Echo Control

Tests indicate that people's tolerance to echo depends upon the loudness of the echo arriving at the talker's ear and the amount the echo is delayed with reference to his direct speech. There is also a relatively wide difference between individuals as to the relative loudness and delays which they will tolerate. The average tolerance to echo obtained from tests under controlled conditions are listed in Table I.

These values form the basis of present designs for intertoll trunk losses. Recent tests made to confirm these values indicate that present-day customers may be somewhat less tolerant to talker echo, particularly at the longer delays, than the figures indicate. However, with the use of carrier facilities on all final route trunks the round-trip delay on calls entirely on one regional center area will be less than the maximum of 45 ms (milliseconds) expected under the plan. This will introduce some margins that should permit the continued use of the present design factors.

Echo control is accomplished on all interregional telephone connections by the use of echo suppressors on all regional center-regional center trunks and on the high-usage interregional trunks as required. On switched connections entirely in one regional area the assigned trunk losses must be large enough to control echo without the benefit of an echo suppressor. This is illustrated in Fig. 3 by the switched connection between telephone *A* and telephone *B*. It is an interesting fact that from an echo standpoint the echo paths returning from the most remote points are the most important since they have the longest delays (round-trip transit time). The various talker echo paths in this connection for customer at telephone *A* together with their round-trip delays are shown in Fig. 4. They range from 1.9 ms for path *A* to 25.9 ms for path *D*. Path *D* is the controlling path if the return losses of the adjacent primary center (PC) meet the objectives outlined later in this paper. This illustrates again that the transmission considerations relating to nation-wide switching must be based on coordinated teamwork since the terminal return losses at telephone *B* have a considerable effect on the talker echo of the customer at telephone *A* and vice versa.

Table I. Average Tolerance to Echo

Round-Trip Delay, Ms	Average Tolerance to Echo-Loss Required in Echo Path, db
0	1.4
20	11.1
40	17.7
60	22.7
80	27.2
100	30.9

An echo suppressor is a voice-operated device that can be inserted in a 4-wire trunk circuit to control echoes on a telephone connection. For example, when the customer at telephone C (Fig. 3) talks, his speech power operates a relay chain in the suppressor which introduces a loss of about 40 db in the oppositely directed transmission path. The loss is under control of a differential balance arrangement so that each customer can break in and interrupt the speech of the other if he raises his voice as to have his speech power about 5 db higher than that of the other customer at the echo suppressor input. The loss must remain in the transmission path long enough after he pauses in his speech to attenuate his return speech (echo) that is reflected across the 2-wire branch of the 4-wire terminating set of the intertoll trunk at the toll center (TC) of customer *A*. This is accomplished by the use of adjustable slow-release relays. The length of time this loss remains in the opposite transmission path is called echo suppressor "hangover" time. In the connection shown this is approximately 57 ms. In the connection shown the echo suppressor is located at the regional center nearest customer *C* and when the customer at *A* talks the hangover can be as low as 12 ms. However, the hangover adjustment on each echo suppressor must be long enough to attenuate the talker echo on any connec-

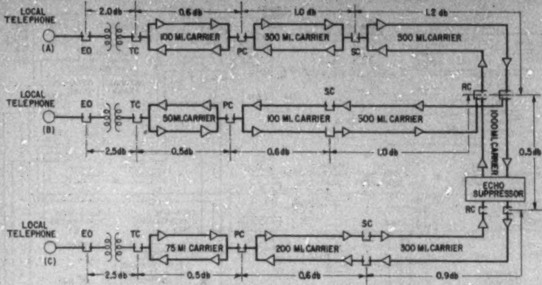

LOCAL TELEPHONE
(A)
EO
TC
PC
SC
2.0db
0.6db
1.0 db
1.2 db
100 MI. CARRIER
300 MI. CARRIER
500 MI. CARRIER
LOCAL TELEPHONE
(B)
EO
TC
PC
SC
RC
50MI.CARRIER
100 MI. CARRIER
500 MI. CARRIER
2.5db
0.5db
0.6db
1.0db
1000 MI. CARRIER
0.5db
ECHO SUPPRESSOR
RC
LOCAL TELEPHONE
(C)
EO
TC
PC
SC
75 MI CARRIER
200 MI. CARRIER
300 MI. CARRIER
2.5db
0.5db
0.6db
0.9db

Fig. 3. Typical losses of various trunks in connections of switching systems

tion to which it may be switched. Tests indicate that best results in the operation of echo suppressors are obtained if the hangover adjustments do not exceed a maximum of 70 ms. For the maximum geographical distances involved in the switching layouts of the United States and Canada this can be accomplished by employing an echo suppressor on each end of the longer RC-RC circuits and interregional high-usage circuits. When two echo suppressors are employed in this manner the suppression circuit in one direction of each echo suppressor is disabled. This permits a reduction in hangover adjustment equal to the round-trip delay (milliseconds) of the RC RC trunk involved.

Singing or Near Singing

In the built-up switched connection of Fig. 3 between telephone *A* and telephone *C*, the echo suppressor will control the echoes. It can, however, be unstable from a singing or near-singing standpoint. Singing can occur around any two 2-wire paths in the connection. Singing occurs if at any one frequency the losses around the singing path become less than the gains. Tests indicate that singing will most generally occur at the low or high frequencies of the voice-frequency band (300 to 3,000 cycles). Tests also indicate that built-up connections become unstable about 6 to 10 db before singing actually occurs, particularly if the instability occurs at frequencies in the lower part of the echo band. In talking over such a connection it sounds like talking into an empty barrel and is often identified as "hollowness."

In order to control both echo and singing it is necessary to set up certain objectives for return losses at the toll centers and at the 2-wire switching offices. It has generally been assumed that the frequencies important from a singing standpoint extended from 200 to 3,000 cycles. This was generally true for voice-frequency intertoll trunks. Recent tests have shown, however, that for carrier facilities the channel band-pass filters tend to introduce more loss at 200 cycles than at 1,000 cycles. This effect makes singing more likely to occur in the frequency range of 300 to 3,000 cycles on switched intertoll connections. It will be noted from Fig. 3 that there are no singing or echo return paths at the 4-wire switching offices. All the losses and return losses can be represented by averages and their variability from these averages or standard deviation. In order to assure freedom from objectionable echo and stability, from singing or near singing on a very high percentage of switched connections, objectives have been established as given in Table II.

Table II. Objectives for Return Losses

	Terminal Toll Offices		2-Wire Switching Offices	
Return Losses	Avg, db	Standard Deviation, db	Avg, db	Standard Deviation, db
Echo (500 to 2500 cycles)	15	3	27	3
Singing (300 to 3000 cycles)	10	2	20	3

The various potential singing paths in the interregional connection shown in Fig. 3 are outlined in Fig. 5. Here again, as mentioned previously for the echo paths, it is important that the return losses in the singing frequency range meet the objectives. If they do, the controlling singing path will most probably be path *A* since the terminal return losses are lower than those on switched connections at the intermediate 2-wire switching offices.

Intertoll Trunk Losses: Deviations From Assigned Losses

Fig. 3 shows the assigned losses of typical lengths of carrier channels that might be involved in a switched connection comprised of final group trunks. Routine tests indicate that these circuit losses vary from their assigned values. A number of factors contribute to the deviations, such as imperfect regulation, error in original line-up, etc. The design factors on which these assigned losses are based assumed that by routine measurements and adjustments together with improved maintenance techniques, the one-way transmission variation could be held to a standard deviation not exceeding 1.0 db. Measurements also indicate that at least for some intertoll trunks the round-trip standard deviation is nearer twice than the 1.73 times the one-way deviation previously assumed. These losses and their

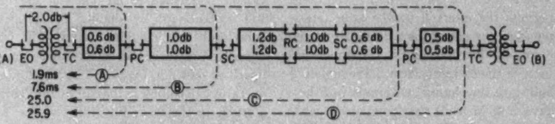

2.0db
0.6db
0.6db
1.0db
1.0db
1.2db
1.2db
RC
1.0db
1.0db
SC
0.6 db
0.6 db
0.5db
0.5db
(A) EO
TC
PC
SC
PC
TC
EO (B)
1.9ms
7.6ms
25.0
25.9
A
B
C
D

Fig. 4. Echo path in connection within regional area

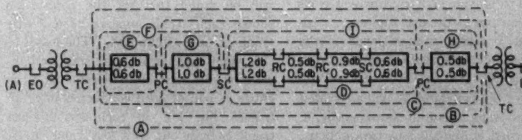

F
E
G
I
H
0.6db
0.6db
1.0 db
1.0 db
1.2db
1.2db
RC
0.5db
0.5db
RC
0.9db
0.9db
SC
0.6db
0.6db
0.5db
0.5db
(A) EO
TC
PC
SC
PC
EO (C)
TC
D
C
B
A

Fig. 5. Singing paths in interregional connection

variations are very important in the results to be obtained in the nation-wide switching of intertoll trunks. On the basis of the losses indicated in Fig. 3 and the distributions, it will be clear that the individual links will be operating at gains instead of the losses indicated a considerable percentage of the time.

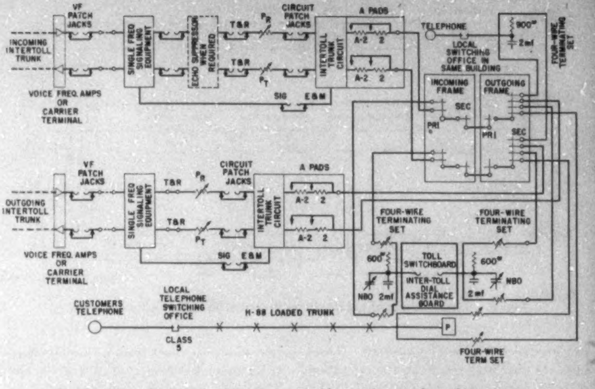


Fig. 6. Through and terminating connection for no. 4 type of switching system

Switching Systems

This paper discusses the transmission features of switching systems employed in the Bell System for nation-wide switching of intertoll trunks. There are other switching systems employed by the independent telephone companies not covered in this paper.[7]

FOUR-WIRE SWITCHING SYSTEM

Fig. 6 shows a schematic arrangement of trunk circuits switched through a 4-wire switching system of the no. *4* type. The sketch is shown on a line-per-pair basis. The 4-wire systems are located at the larger switching offices. On a through intertoll-to-intertoll switch there are no echo or singing paths in the switching office. Most of the intertoll trunks now consist of carrier channels which are of the 4-wire type (a separate channel for each direction of transmission). This eliminates the need for 4-wire terminating sets for intertoll-to-intertoll connections. The carrier system terminates in the voice-frequency (VF) patch bay for testing access to the carrier system independent of the switching system. Both intertoll trunks shown are equipped with a single-frequency signaling system employing a frequency in the voice-frequency band that can readily be transmitted over any of the message channels. It will be noted that it terminates in the trunk circuit by means of E & M signaling leads. This is standard for the Bell System and permits the intertoll trunks to be terminated in any one of the switching systems discussed in this paper. Channels of some of the carrier systems have signaling built into the channel terminals and they have E & M leads that co-ordinate with the trunk circuits associated with the various switching systems.

Carrier channels are inherently arranged to provide high transmission levels at the output of the demodulators. Advantage is taken of these levels in the 4-wire switching system to provide gain as required. This is accomplished by means of the switchable *A* pad associated with the trunk circuit as shown in Fig. 6. The *A* pad is divided into two sections. One section has a loss of 2 db and the other (*A*-2) db. In switching systems currently installed the (*A*-2) section is usually 5 db thereby making the total *A* pad 7 db. Each section is controlled by contacts of separate relays. As shown in Fig. 6, the entire *A* pad remains in the connection on each intertoll trunk on an intertoll-to-intertoll switch. The *PR* pad is adjustable in steps of 0.25 db to adjust the loss of each intertoll trunk to provide the specified loss from the switches in the remote office to the switches in the office shown in Fig. 6. This is known as via net loss (VNL) and is the lowest loss at which the trunk can be operated, limited by its inherent characteristic and the variations in return losses at the terminals and at 2-wire switching points. The *PT* pad is used to provide the proper level at the input of the various carrier or voice-frequency systems.

If an operator at some distant office desires assistance in completing a call she can be connected to an operator at the toll switchboard by dialing the proper code and the incoming trunk will be connected to the trunk to the toll switchboard as shown in Fig. 6. The assistance operator can complete the call by selecting a trunk to the incoming frame of the 4-wire machine which will connect to any outgoing trunk to the intended telephone upon receipt of the proper dial code. It will be noted that for the connection shown each trunk appears as a 2-wire trunk at the switchboard. The losses of the two 4-wire terminating sets are canceled by gain introduced by removal of the entire *A* pad of each connected intertoll trunk. The loss in each switchboard trunk is built out to a loss value of *A*. On the switch the loss is equal to $-A+A-A+A=0$. It will be seen that the connection through the switchboard does not introduce a loss penalty to the connection. It will be necessary, however, to build out the capacitance in both 2-wire switchboard paths to a uniform value and adjust the network building out (BO) capacitors across the compromise network of each 4-wire terminating set to the correct value based on return loss measurements. Two-wire switchboards are generally employed, since the operators can complete other calls when not occupied with assistance calls. At some of the 4-wire switching offices the assistance switchboard is of the 4-wire type with 4-wire cord circuits.

When the incoming intertoll trunk is connected to a switching trunk to the local office it connects to the 4-wire side of a 4-wire terminating set associated with the 2-wire loaded trunk to the local office. Since the 4-wire terminating set is connected directly to the loaded trunk a precision network can be used to match the loaded trunk and high terminal balance can be obtained so that gain can be introduced into the trunk by removing the entire *A* pad (7 db). Four decibels of this gain cancel the loss of the 4-wire terminating set and 3 db are available to offset some of the loss in the local trunk. In the connection shown in Fig. 6 the 24-gage *H88* loaded switching trunk to the remote class *5* office has a trunk loss of 6 db. It is connected to the switching office outgoing frame by means of a combined 4-wire terminating set and trunk circuit. It is arranged with a simplex mark so that when seized by an incoming intertoll trunk the entire *A* pad is removed and 3 db of the loss of the trunk are offset by the gain obtained from the *A* pad removal.

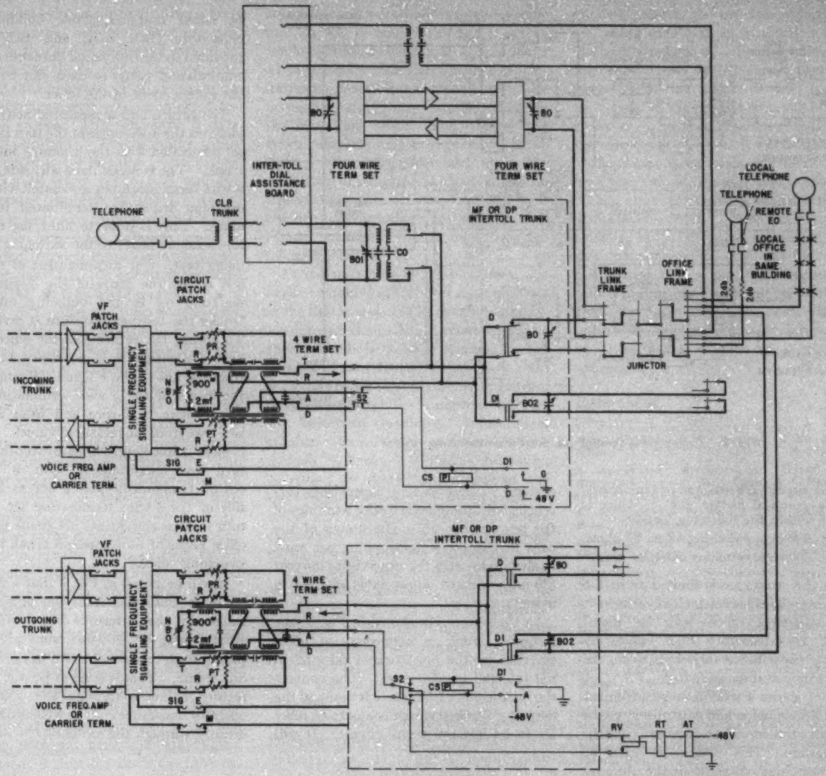

INTER-TOLL DIAL ASSISTANCE BOARD
FOUR WIRE TERM SET
FOUR WIRE TERM SET
BO
BO
LOCAL TELEPHONE
TELEPHONE
REMOTE EO
LOCAL OFFICE IN SAME BUILDING
TELEPHONE
CLR TRUNK
MF OR DP INTERTOLL TRUNK
BO1
CO
TRUNK LINK FRAME
OFFICE LINK FRAME
2db
2db
JUNCTOR
CIRCUIT PATCH JACKS
VF PATCH JACKS
INCOMING TRUNK
SINGLE FREQUENCY SIGNALING EQUIPMENT
T
R
PR
PT
N
B
O
900Ω
2 mf
4 WIRE TERM SET
T
R
A
D
S2
D
D1
BO
BO2
SIG
E
M
VOICE FREQ. AMP OR CARRIER TERM.
CS
P
D1
G
D
48 V
CIRCUIT PATCH JACKS
MF OR DP INTERTOLL TRUNK
VF PATCH JACKS
OUTGOING TRUNK
4 WIRE TERM SET
BO
BO2
S2
CS
P
D1
A
-48V
VOICE FREQ. AMP OR CARRIER TERM.
RV
RT
AT
-48V

Fig. 7. Dial-to-dial intertoll connection through crossbar tandem showing supervisory and transmission paths

The net trunk loss is now 3 db instead of 6 db. If the switching trunk is terminated on a local switching machine in the same building as the 4-wire switching system only the 2-db section of the *A* pad is removed to cancel 2 db of the 4-db loss in the 4-wire terminating set. The 2-db remaining loss in the 4-wire terminating set acts as a 2-db equivalent pad required for return loss improvement.

Two-Wire Switching Systems

The economics of the no. *4* type of switching system favor its use in the larger switching offices. It is somewhat more expensive than 2-wire switching systems at those offices where the number of intertoll trunk terminations do not exceed about 3,000. Consequently, 2-wire switching systems are employed at the intermediate-sized and smaller-sized switching offices.

When 2-wire switching systems are employed, singing and echo paths are introduced in the switching office. Since any trunk must switch to any other, the most practicable way to handle the balance problem is to make all circuits uniform in impedance to permit a single value of compromise network. It is important, therefore, to give careful attention to the design of the various equipment components that are employed in the transmission paths. The following transmission requirements should be met if satisfactory results are to be obtained at 2-wire switching offices:

1. The 4-wire terminating sets must have a flat transmission-frequency characteristic in the useful transmitted voice band (300 to 3,200 cycles) in order to avoid undue transmission distortion in a switched connection. Two such 4-wire terminating sets are involved in an intertoll-to-intertoll switch at each 2-wire switching office.

2. Four-wire terminating sets must have an impedance (viewed from the 2-wire side) that closely approximates the compromise balancing network selected for the office.

3. All equipment and wiring used to establish the connection must make only a very small contribution to the impedance of the connection, for instance:

(*a*) All bridged impedances (across the transmission paths) must have sufficiently high impedance over the important voice-frequency band so that they do not appreciably effect the office return loss.

(*b*) Switching office wiring layouts should be carefully planned to keep the effect of the maximum series resistance of the wiring in the office path when two intertoll

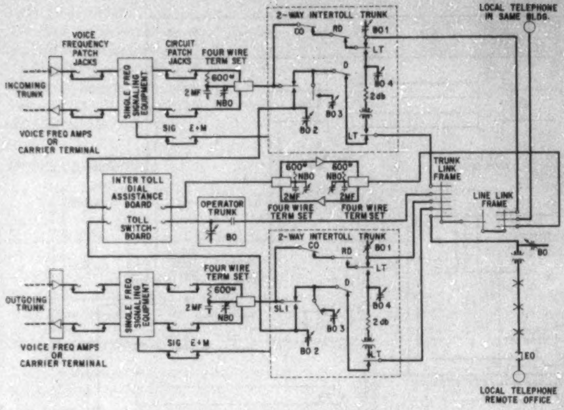

2-WAY INTERTOLL TRUNK
LOCAL TELEPHONE IN SAME BLDG.
VOICE FREQUENCY PATCH JACKS
CIRCUIT PATCH JACKS
FOUR WIRE TERM SET
CO
RD
BO 1
LT
D
BO 4
2db
600ʷ
2 MF
NBO
INCOMING TRUNK
SINGLE FREQ SIGNALING EQUIPMENT
BO 3
BO 2
LT
SIG
E+M
VOICE FREQ AMPS OR CARRIER TERMINAL
TRUNK LINK FRAME
600ʷ
NBO
600ʷ
NBO
INTER TOLL DIAL ASSISTANCE BOARD
2MF
2MF
FOUR WIRE TERM SET
FOUR WIRE TERM SET
LINE LINK FRAME
OPERATOR TRUNK
TOLL SWITCH-BOARD
BO
2-WAY INTERTOLL TRUNK
CO
RD
BO 1
BO
LT
FOUR WIRE TERM SET
600ʷ
D
OUTGOING TRUNK
2 MF
NBO
SL1
BO 4
2 db
BO 3
SIG
E+M
VOICE FREQ AMPS OR CARRIER TERMINAL
BO 2
LT
EO
LOCAL TELEPHONE REMOTE OFFICE

Fig. 8. Connections through no. 5 crossbar switching system

trunk circuits are switched to a minimum. As a practical matter, the resistance is kept within the following values:

(1). 900-ohm switching office, 65 ohms.
(2). 600-ohm switching office, 45 ohms.

When the resistance is limited as in the foregoing, the office return loss paths in 2-wire switching offices can be maintained at satisfactory high values by making capacitance adjustments in the office wiring as discussed later.

Fig. 7 shows a schematic arrangement of the transmission and supervisory paths through the crossbar tandem switching system. This system is used at those switching offices where the ultimate number of trunk terminations is not expected to exceed its capacity. Fig. 7 is drawn on a line-per-wire basis in order to show more clearly the supervisory paths through the office. The crossbar tandem employs reverse battery loop supervision through the office and consequently the supervisory relay (CS) is bridged across a capacitor in the transmission path. This relay must, therefore, be designed to stand the supervisory current through its winding without appreciable change in its inductance. The crossbar tandem system does not employ a trunk circuit with repeating coils for impedance matching in the switching trunks to the local offices. As a result, the offices' impedance has been designed to be 900 ohms. Therefore, the 4-wire terminating sets associated with the intertoll trunks are designed to match the 600-ohm impedance of the carrier channels or voice-frequency amplifiers to the 900-ohm impedance of the 2-wire switching system. The office loop supervisory current is fed through the 2-wire windings of the terminating set. The design of the coils in the 4-wire terminating set must be such as to carry the supervisory current (30-mil) without appreciable change in inductance.

In the connection shown in Fig. 7, the incoming trunk is connected to office switches by the operation of relay *D* in the intertoll trunk circuit. The connection at the office link frame is made to the outgoing trunk through contacts of relay *D1* in its intertoll trunk circuit. It will be noted that adjustable building-out capacitors (BO, BO*1*, and BO*2*) are provided in the individual branches of the transmission paths so each can be built out independent of the other.

The return loss is measured across the jacks on the 4-wire side of the terminating set associated with the incoming intertoll trunk. The jacks on the 4-wire side of the 4-wire terminating set associated with the outgoing trunk are terminated in 600 ohms. This is done to limit the return loss measurements to the switching office under test. For measurement of return loss across the jacks of the 4-wire terminating set in the outgoing intertoll trunks the testing procedure is reversed. To obtain the return loss for office balance evaluation it is necessary to subtract the loss of the *PT* and *PR* pads and the losses of the 4-wire terminating set from 4-wire to 2-wire and from 2-wire to 4-wire from the measured values. This is most conveniently done by another return loss measurement with a short circuit temporarily placed across 2-wire side of the 4-wire terminating set. Return loss measurements are made on the office paths of each intertoll trunk in the switching office.

It will be noted in Fig. 7 that a 2-wire trunk is employed from the office link frame to the intertoll dial assistance switchboard. A 4-wire trunk is employed from this switchboard to the trunk link frame. This is done to permit satisfactory balances through the office on operator assistance calls as well as on calls directly through the switches.

Vol, Pt. 1 195

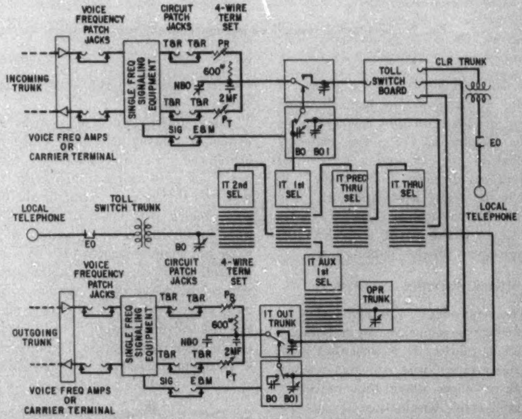

VOICE FREQUENCY PATCH JACKS
CIRCUIT PATCH JACKS
4-WIRE TERM SET
T&R
T&R
P_R
600ʷ
INCOMING TRUNK
SINGLE FREQ SIGNALING EQUIPMENT
NBO
2MF
T&R
T&R
P_T
SIG
E&M
VOICE FREQ AMPS OR CARRIER TERMINAL
TOLL SWITCH BOARD
CLR TRUNK
BO
BO1
EO
IT 2nd SEL
IT 1st SEL
IT PREC THRU SEL
IT THRU SEL
LOCAL TELEPHONE
TOLL SWITCH TRUNK
LOCAL TELEPHONE
EO
BO
IT AUX 1st SEL
VOICE FREQUENCY PATCH JACKS
CIRCUIT PATCH JACKS
4-WIRE TERM SET
T&R
T&R
P_R
OPR TRUNK
600ʷ
IT OUT TRUNK
OUTGOING TRUNK
SINGLE FREQ SIGNALING EQUIPMENT
NBO
T&R
T&R
2MF
P_T
SIG
E&M
VOICE FREQ AMPS OR CARRIER TERMINAL
BO
BO1

Fig. 9. Through and terminating connections for step-by-step switching system

At some of the smaller switching offices where a single switching machine can handle both the toll and local switching requirements, the no. *5* crossbar switching system is employed. Fig. 8 shows a schematic arrangement of an intertoll-to-intertoll switch. This schematic is on a line-per pair basis. The incoming trunk is connected to the line link frame by the operation of relays (*CO*) (*RD*) and the nonoperated contacts of the *LT* relay. The switches close through to the trunk link frame where the outgoing trunk is connected through the nonoperated contacts of relay *LT*, the operated contacts of relay *RD* and the nonoperated contacts of relay *SLI*. The no. *5* crossbar system employed for toll switching has been designed for an office impedance of 600 ohms. The 4-wire terminating sets employed at this office have a 1-to-1 impedance ratio from 4-wire to the 2-wire side. The compromise balancing network consists of a 600-ohm resistance in series with a 2-microfarad capacitor. The various branches of the 2-way intertoll trunk circuits are also equipped with adjustable capacitors (BO*1*, etc.) to permit office paths capacitance adjustments to provide for the necessary high return loss similar to that previously discussed in connection with the crossbar tandem switching system.

At some of the switching offices, particularly primary centers, the step-by-step switching system was already in operation when the present nation-wide switching plan was devised. Since the step-by-step system employed in the Bell System does not have common control equipment it does not have the inherent capabilities of selecting high-usage trunks for alternate routing. However, from a transmission standpoint, the office echo and singing paths can be controlled by compact office layouts and office wiring balance work as discussed for the previously mentioned 2-wire switching systems. Fig. 9 shows an intertoll-to-intertoll connection through the step-by-step switching system. It will be noted that adjustable capacitors are provided in the various transmission paths through the office to permit the building out of the capacitance of the office cabling to uniform values as required. The nominal impedance of the step-by-step switching system is 600 ohms. The 4-wire terminating sets associated with the intertoll trunk circuits are of the type having an impedance of 600 ohms on both the 4-wire and 2-wire branches. The compromise networks are composed of a 600-ohm resistance in series with a 2-microfarad capacitor and are equipped with an adjustable network-building-out capacitor. It will be noted from Fig. 9 that it employs trunk circuits with E & M leads which connect with the standard single-frequency signaling system to transfer the necessary supervisory signals through the switching office. The step-by-step system employs simplex supervision paths through the office obtained from the center tap of inductors bridged across the transmission paths in the intertoll trunk circuits. Their inductance must be high enough so as not to alter materially the office impedance.

In the next 5 or 10 years most of the major switching offices now contemplated in the Bell System will be in and operating. Equipment is also being provided which should result in a large increase in the number of calls dialed by customers. On such calls where any number of links from one to nine or ten are connected together by switching machines, transmission troubles on a particular connection will be difficult to locate. It will be necessary, therefore, to design and maintain the individual components by means of initial and periodic tests that will insure satisfactory performance on a very high percentage of switched connections.

Only about half of the ultimate 300 major machine switching offices are now in operation. It is planned to obtain information on their performance by means of transmission and return loss tests. This information will be helpful in the design and operation of systems yet to be installed.

A fairly high percentage of intertoll trunks are now operating at the low losses indicated in Fig. 3. A considerable number, meeting the new loss objectives, are added each year in connection with the addition of new switching offices or rearrangements. It is believed that the losses shown are about as low as can be reached in the present state of the art. However, if further experience indicates that they can be lowered it will be done.

The transmission engineer's dream is to have zero loss on all connections regardless of the number of links in tandem or distance involved. This cannot be achieved in the immediate future but with new instrumentalities and new developments it should be possible to approach nearer and nearer to the ultimate objective. In the meantime, he is striving to operate all existing plants to insure the best possible transmission on all switched connections.

References

1. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, H. R. Huntley. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 670–76.

2. AUTOMATIC SWITCHING FOR NATION-WIDE TELEPHONE SERVICE, A. B. Clark, H. S. Osborne. *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 245–48.

3. FUNDAMENTAL PLANS FOR TOLL TELEPHONE PLANT, J. J. Pilliod. *Ibid.*, pp. 248–56.

4. NATION-WIDE NUMBERING PLAN, W. H. Nunn. *Ibid.*, pp. 257–60.

5. AUTOMATIC TOLL SWITCHING SYSTEM, F. F. Shipley. *Ibid.*, pp. 261–69.

6. IMPEDANCE AND RETURN LOSS PERFORMANCE OF TELEPHONE PLANT IN METROPOLITAN AREAS, L. B. Bogan. *Ibid.*, vol. 77, pt. I, 1958 (*Paper no. 58-4*).

7. RECENT DEVELOPMENTS IN 4-WIRE SWITCHING OF LONG-DISTANCE TELEPHONE CIRCUITS, Imre Molnar. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 722–33.

8. INTERTOLL TRUNK NET LOSS MAINTENANCE UNDER OPERATOR DISTANCE AND DIRECT DISTANCE DIALING, H. H. Felder, E. N. Little. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1946, p. 955.

Discussion

Arthur F. Rose (American Telephone and Telegraph Company, New York, N. Y.): Mr. Wallman's paper and the related papers of Bogan[1] and Speake[2] must be considered as a whole to appreciate the variety of problems involved in giving good nation-wide toll telephone service.

Mr. Wallman's paper emphasizes the problems at the intermediate switching points where it is desirable to eliminate in so far as practicable the echo and singing paths. If economics would permit, all intermediate points would be switched on a 4-wire basis and this problem would disappear, but the practical conclusion is that we cannot afford this luxury. Many of the smaller offices will continue on a 2-wire basis indefinitely, and Mr. Wallman's paper shows how their echo contributions can be made negligible by proper design and engineering procedures.

However, the point I wish to emphasize is that it is not enough to do this initially, but eternal vigilance is required to insure it remaining so. Every time new switchboard positions or multiple are added at a 2-wire switching point, checks should be made and building-out capacitors adjusted if necessary to avoid the gradual erosion of the initial high office-return loss.

The maintenance costs involved in this continual checking offset to some extent the first-cost advantage associated with 2-wire switching in the smaller offices. However, the practical approach is not to try to have our cake and eat it too, but to carry out honestly the commitments which we tacitly assumed when we elected to use 2-wire switching as the proper engineering solution in a particular switching center. As time goes on and new plant personnel take over the office, it will be easy to forget the need for spending this maintenance money in those cases where first-cost savings may have been substantial.

In an unpublished paper by J. N. Petrie it was claimed that the well-known "tail-eating" scheme avoids the balance worries inherent in 2-wire switching systems described by Mr. Wallman and is simpler than the full 4-wire treatment used in the Bell System. From a practical standpoint again, continuing maintenance attention should be balanced against first cost savings and permanent good results will not be attained without checks from time to time.

REFERENCES

1. See reference 6 of the paper.

2. STUDY OF RETURN LOSSES ON LOADED TRUNK CABLES AND METHODS TAKEN TO IMPROVE THEM, G. H. Speake. *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 201–04.

E. A. Bourne (Bell Telephone Company of Canada, Montreal, Que., Canada): As Mr. Wallman has pointed out, co-ordinated teamwork among operating companies and suppliers is necessary for satisfactory performance on toll connections. The aim of this discussion is to enlarge on this aspect, drawing upon the Canadian switching network to illustrate the need.

In Canada, our major toll switching centers utilize equipment manufactured by several suppliers and are operated and maintained by various fully independent organizations. For example, of nine of our largest cities two are equipped with Bell System *4A* crossbar, two will use General Telephone *FW-1* equipment, three have Siemens auto toll equipment, and two are being equipped with Bell System crossbar tandem. Eight are major switching centers and are operated by six independent organizations. Interconnection of these systems could result in compromised transmission objectives if any system is permitted to introduce appreciable echo or singing paths. In the case of echo, the results of poor return loss are not apparent to the customer at the point of poor return loss, but indeed to the customer at the other end of the connection, perhaps 1,000 miles away. It is evident that the long-distance customer is not the exclusive responsibility of one company but the joint responsibility of all and it is for this reason that the adoption of uniform transmission standards by suppliers and operators is necessary.

Mr. Wallman has dealt mainly with switching equipment in use in the Bell System. Another type of switching system, very much on the Canadian scene is the "tail-eater," a type of 4-wire switching equipment manufactured by the Automatic Electric Company and by Siemens Brothers of England. The transmission performance of the Automatic Electric *FW-1* system has been covered by Dr. Molnar (see reference 7 of the paper). Siemens equipment has been tested from an office balance standpoint by the Manitoba Telephone System and results indicate good performance in echo and singing ranges.

The Application of Negative-Impedance Repeaters on Long Rural Telephone Lines

H. T. UTHLAUT, JR.
MEMBER AIEE

PRESENT-DAY expansion of telephone service is at peak levels with no immediate end in sight, and benefiting by this expansion is an important segment of the United States—the rural area. Questions may be asked, "What's new about this?" or, "Rural telephones have been with us for years, so why the fuss?" True, there is nothing new in the concept of rural telephone service, but the present demand is at such levels that serious problems are presented to the telephone industry. The economic problem is one of real significance, but it is not discussed here. It is proposed to discuss, rather, the transmission aspects involved in extending this service to these areas.

Problem

Studies have shown that there is a great demand in rural areas where telephone service has not heretofore been provided. In those areas where the demand has been small, open-wire circuits have generally been adequate and satisfactory, both from a desirable transmission viewpoint and from a construction standpoint. The present large demand, however, dictates the use of cable facilities to replace the inadequate open wire, and requires the loops to extend to lengths unprecedented in normal exchange service. It is not unusual in present engineering of rural telephone service to encounter conductor loops well in excess of 100,000 feet from the serving telephone central office. When, as so often is the case, the demand is too small to establish economically a new central office and too great to allow open wire to be used, cable plant, with its resultant high attenuation, must be proposed.

New Types of Plant

The ever-increasing search for newer and better products is in "full swing" in all industry in the United States, and the telephone industry is no exception. In addition to the many electronic devices under development, new outside plant is also in evidence. Some of the more recently developed items include polyethylene-insulated conductor cable,[1] *B* rural wire (polyethylene-insulated pairs spiralled about a steel support wire), *B* underground wire (rubber-insulated pair of no. 16 American Wire Gauge copper conductors with a steel-armor and a neoprene jacket), and *C* rural wire (one pair of polyethylene-insulated copper-steel conductors suitable for extra-long-span construction). Because of the high dielectric strength of the insulating material, it is expected that troubles on these types of plant will be less than that now experienced on paper- and pulp-insulated conductor cable.

These new types of outside plant are answering many of the needs of the telephone engineer, and are being proposed in a great many of the rural projects now being engineered.

Transmission Limitations

Having been apprised of types of suitable plant, the engineer may now see that the demand for telephone service must needs be satisfied by use of cable plant in an appropriate size, ranging from as little as six pairs to as many as 101 pairs in most projects, and possibly even greater in others. From the commercial viewpoint, it is seen that the demand will be served by cable. From the viewpoint of the outside plant engineer, the new types of cable plant, for both aerial and buried use, seem to be satisfactory. In contrast, the transmission engineer views with concern the proposal of voice-frequency subscriber loops of 90,000 feet of cable plant and 20,000 feet of open wire extending beyond that. It is in this area of telephone engineering where the present resistance-design philosophy is not adequate.[2] The only practicable method of meeting the d-c resistance limits of the central office

---

Paper 57-391, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted January 3, 1957; made available for printing May 13, 1957.

H. T. UTHLAUT, JR., is with Southern Bell Telephone and Telegraph Company, Birmingham, Ala.

in these cases is normally to use the coarsest gage cable in general use, which in the telephone industry is no. 19 American Wire Gauge. Therefore, the transmission engineering design has been reduced to calculation of the proposed subscriber loop to determine if objectives are exceeded.

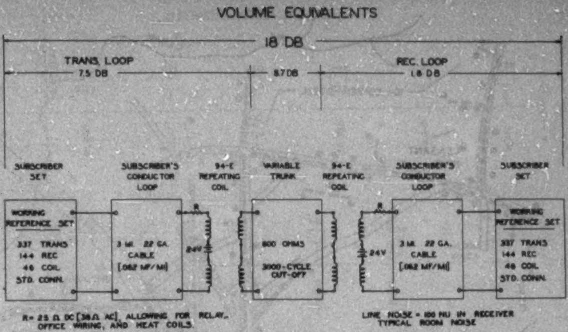


Fig. 1. Working reference system for specifying effective losses; equipment shown is of Western Electric nomenclature

When calculations of effective losses are made on these long subscriber loops, it is found that the use of cable plant of approximately 85,000 feet and open-wire extensions beyond this figure generally indicate that the loop, when compared with the working reference system described in Fig. 1, is about 0 db (decibels) at the end of the cable, if the cable is fully loaded at *H* spacing (6,000-foot intervals) with 88-millihenry loading coils, and the station sets use three cells of local battery and *HC5* receiver.

In this era of ever seeking to decrease the transmission losses on exchange subscriber loops (as well as on trunks), it is well to examine the transmission engineer's problem here. It has been his desire to establish the goal of maximum subscriber loops of −3 to 0 db (compared with the working reference system), and the closer to −3 db the better. Therefore, with the rural program being developed to a greater extent each day, the telephone engineers must maintain the high standard of transmission design to meet these established transmission goals. Cost studies prove that it is generally unsound economical practice to use coarse-gauge cable to reduce transmission losses. A better method of meeting these goals is by the use of *E*-type repeaters on these exchange loops. These repeaters are negative-impedance devices which are relatively inexpensive to install. They require no expensive filters or bypass signalling equipment. More information on this equipment is available in other literature.[3–6]

Examples

A typical example of an engineering design on which it was found necessary to provide for the use of repeaters is presented in the following.

Fig. 2 is a map of the area involved, showing the telephone demand and type of plant proposed. An estimate of the area's growth indicates that line requirements as shown on the map are expected. On the basis of the line requirements, the outside-plant engineer must propose the plant which is the economic choice. This plant is indicated on the map. Existing open-wire plant is also shown, and almost all of the existing feeder open-wire plant is proposed to be replaced by cable-feeder plant. The subscribers served by existing plant are not shown on the map, but an indication of the existing requirements, which are seven lines, is shown at the end of the cable.

Fig. 3 is a schematic diagram which shows the specific design of the facilities proposed to serve the demand in the area. It is this schematic which the transmission engineer will examine closely to assure that the design meets the prevailing objectives. In actual practice, more information is added to this schematic, such as pairs being distributed beyond the load points, and pairs cut dead to prevent excessive multiple beyond the location at which they are distributed.

The following items are included in the examination of the design:

1. Resistance check for supervision limits as shown in Table I. With use of proper types of central-office equipment to extend signaling ranges, this portion of the design is satisfactory.

2. Load-coil spacing on *H* design, as shown in Table II. Deviation of the average spacing from standard (6,000-foot) spacing is 10 feet or 0.2%.

This design meets a reasonable objective in subscriber-loop loading; i.e., average spacing within 2% of the standard 6,000-foot spacing, individual sections within 3% of the average of the design, and average deviation within 1.2% of average section.

3. Distribution of fully loaded pairs in proper zones. Combined with this part of the design is the use of a reasonable amount of multiple tap, taking into consideration both the transmission penalty incurred by excessive amounts and the need to increase it to an amount suitable for obtaining a satisfactory percentage of 8-party stations per line.

4. Determine the effective loop loss (compared with the working reference system) using the data shown in Table III for local battery station equipment, as follows:

a. Losses
   $(3)/(1) = 1.3$ db
   $(2) = 30.7$ db
   Miscellaneous losses $= 0.5$ db (loss for transformers, drop, etc.)
      Total losses—32.5 db
b. Gains
   Loop-loading gain $= 1.5N - 1.5$, for 22, 24, 26 gage (N = no. of load coils)
   Loop-loading gain $= 1.1N - 1.5$, for 19 gauge
      Total gain $= -17.3$ db
c. Corrections
   Correction for local battery loop $= -11.0$ db
   Correction for more efficient transmitter and receiver, and for three cells of local battery $= -2.0$ db
   Gains are shown as negative and losses as positive.
d. Totals
   Total losses $= +32.5$ db
   Total gains $= -30.3$ db
      Loop loss $= +2.2$ db $(T+R)/2$

In the examination of this design, the transmission engineer has determined that the only unsatisfactory aspect is the transmission loss. The means of reducing the transmission loss have been discussed previously. The method of reducing this loss by replacing existing cable with larger gage is a possibility. It would also be possible to consider use of even a coarser gage, such as 16 gage. These two methods are generally prohibitively costly, however, and another solution should be explored. This leads naturally to consideration of the use of an *E*-type repeater.

Considerations in Repeater Design

All the aspects in the application of the *E*-type repeater are not discussed here. Suffice it to point out that this layout

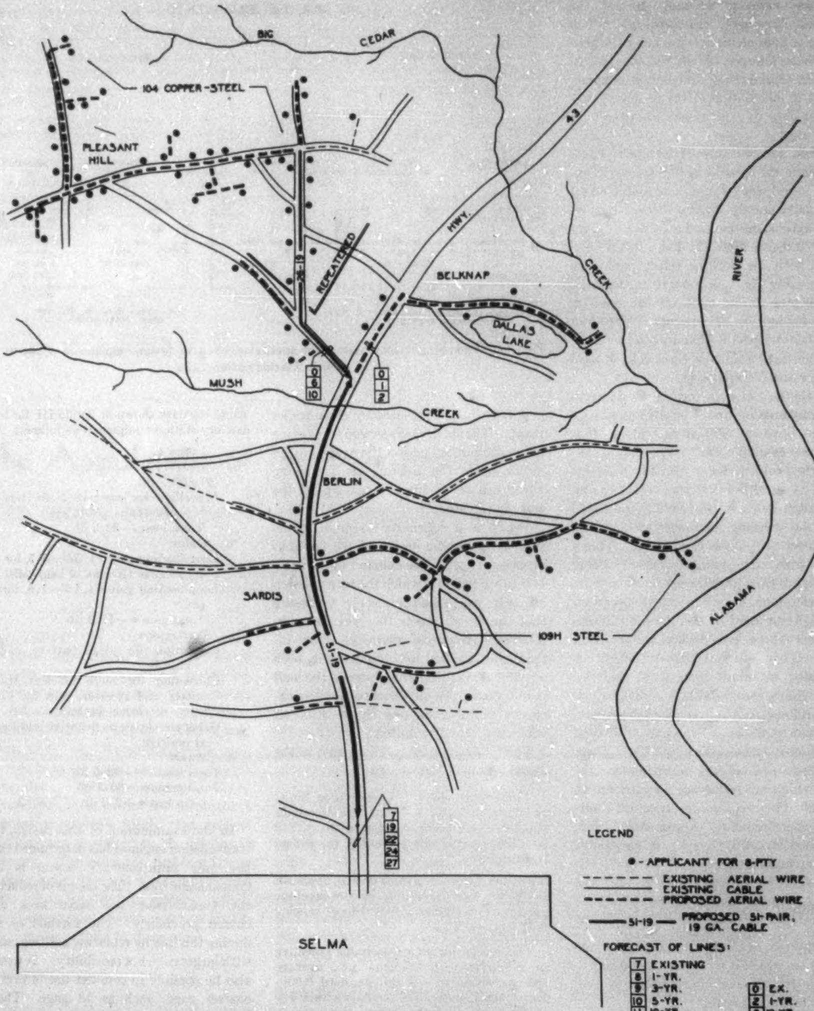

BIG
CEDAR
104 COPPER-STEEL
PLEASANT HILL
43
HWY.
REPEATERED
26-19
BELKNAP
DALLAS LAKE
CREEK
RIVER
MUSH
CREEK
0 6 10
0 1 2
BERLIN
SARDIS
109H STEEL
ALABAMA
51-19
7 19 22 24 27
SELMA
LEGEND
● - APPLICANT FOR 8-PTY
EXISTING AERIAL WIRE
EXISTING CABLE
PROPOSED AERIAL WIRE
51-19 PROPOSED 51-PAIR, 19 GA. CABLE
FORECAST OF LINES:
7 EXISTING
8 1-YR.
9 3-YR.
10 5-YR.
11 10-YR.
0 EX.
2 1-YR.
8 10-YR.

Fig. 2. Area map showing telephone demand and type of plant proposed

lends itself adequately to the use of a repeater. Where there is a well-planned utilization of cable gages and a properly designed loading layout, a high structural return loss may be expected with cable manufactured to stringent tolerances. An *E-2* repeater is chosen over the more expensive *E-23*, because adequate gain to meet present objectives generally may be obtained from the former. To meet future objectives, it may be possible to consider adding the *E-3* element to the proposed *E-2* repeater.

As discussed in the foregoing, the reference point of a 0-db $(T+R)/2$ loop was established as the point beyond which loops should have a repeater placed in the circuit at the serving central office. This could just as arbitrarily be chosen at −1 db or −2 db, which would produce even better transmission on the subscriber-loop plant. There are several limiting factors on this decision, however, and the most difficult one to satisfy is the supply situation. For the next few years there will not be enough repeaters available to hold all subscriber loops as low as −2

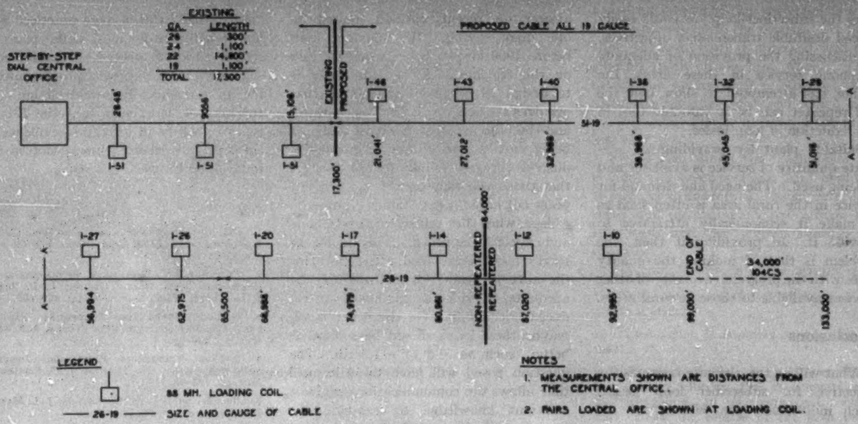

EXISTING
GA. LENGTH
26 300'
24 1,100'
22 14,800'
19 1,100'
TOTAL 17,300'
STEP-BY-STEP DIAL CENTRAL OFFICE
PROPOSED CABLE ALL 19 GAUGE
EXISTING
PROPOSED
2848'
9058'
15,108'
I-51
I-51
I-51
17,300'
I-46
21,041'
I-43
27,012'
I-40
32,988'
51-19
I-36
38,988'
I-32
45,045'
I-26
51,018'
A
A
I-27
56,994'
I-26
62,975'
65,500'
I-20
68,988'
I-17
74,979'
26-19
I-14
80,981'
NON-REPEATERED
REPEATERED
84,000'
I-12
87,020'
I-10
92,985'
END OF CABLE
99,000'
34,000'
104CS
133,000'
LEGEND
88 MH. LOADING COIL
26-19 SIZE AND GAUGE OF CABLE
NOTES
1. MEASUREMENTS SHOWN ARE DISTANCES FROM THE CENTRAL OFFICE.
2. PAIRS LOADED ARE SHOWN AT LOADING COIL.

Fig. 3. Schematic diagram

db. Therefore, it appears that the use of the 0-db point is more realistic. For the example used in the paper, the requirements for repeaters on loops greater than 0 db $(T+R)/2$ have been indicated. This point is located at approximately 84,000 feet from the central office.

For the layout shown in Fig. 3, a repeater gain of 3.5 db may be reasonably expected from design calculation. With this gain, the effective subscriber loop will be -1.3 db $(T+R)/2$ compared with the working reference loop. Therefore, by using an *E-2* repeater in the circuit, the loss is within a realistic transmission design objective, thus rendering satisfactory a subscriber loop which would have been beyond the boundary of good engineering design.

As an example of how some of these considerations find their way into the design of rural telephone plant, there have been nine rural projects in the Alabama division of the Southern Bell Telephone and Telegraph Company which have specified *E-2* repeaters on certain loops. A total of over 110 repeaters have been required on these projects which had effective subscriber loops from $+0.7$ to $+3.6$ db $(T+R)/2$. Nominal gain of *E-2* repeaters on these loops were 3.5 to 4.0 db, with the resulting calculated design of better than 0 db $(T+R)/2$ on all projects. In addition to these projects where definite plans have been made for repeaters, others are under consideration to bring losses within reasonable objectives. No figures are available as to the possible number of repeaters required for these additional areas.

Summary

Today's rural area is much different from that known by our forebears. One thinks fondly of a past era which included the kerosene lantern, wood-fired range, kitchen washstand and, if it were a progressive area, a farmer telephone line, but certainly no one would be willing to exchange the modern electric lamps, electric household appliances, plumbing systems, and adequate telephone service for the antiquated items of yesteryear.

After examination of the problem presented by extension of telephone service into the rural area, it is seen that in many

Table I. Resistance Check for Supervision Limits

Type of Plant	Length, Feet	Ohms
Cable, 26 gage at 68 degrees	300	25
Cable, 24 gage at 68 degrees	1,100	52
Cable, 22 gage at 68 degrees	1,600	52 (existing)
Cable, 22 gage at 100 degrees	13,200	458
Cable, 19 gage at 100 degrees	1,100	18
Cable, 19 gage at 68 degrees	81,700	1,322 (proposed)
16 88-millihenry load coils		144
Wire, 104 copper steel	34,000	170
Totals	133,000	2,241

Table II. Load-Coil Spacing on H Design

Full Sections	Spacing, Feet	Deviation From Average Spacing	
		Feet	Per Cent
1	6,208	+198	+3.3
2	6,052	+ 42	+0.7
3	5,933	− 77	−1.3
4	5,971	− 39	−0.7
5	5,974	− 36	−0.6
6	5,982	− 28	−0.5
7	6,077	+ 67	+1.1
8	5,971	− 39	−0.7
9	5,978	− 32	−0.5
10	5,981	− 29	−0.5
11	6,013	+ 3	0
12	5,991	− 19	−0.3
13	6,002	− 8	−0.1
14	6,039	+ 29	+0.5
15	5,975	− 35	−0.6
Averages	6,010	46	0.8

Table III. Data for Effective Loop Loss

Facility	Length, Kilofeet	Factor	Length Factor	Constant	Length Constant
Cable, 26 gage	0.3	0.578	0.2	2.9	0.9
Cable, 24 gage	1.1	0.437	0.5	2.8	3.1
Cable, 22 gage	14.8	0.329	4.9	1.9	28.2
Cable, 19 gage	82.8	0.288	23.8	0.8	66.3
Wire, 104 copper steel	34.0	0.037	1.3	2.1	71.5
Totals	133.0 (1)		30.7 (2)		170.0 (3)

cases the subscriber-loop loss could easily exceed desirable transmission objectives. In discussing the provision of adequate telephone service in these areas, the author has attempted to show how the *E-2* repeater can be a valuable tool in the reduction of loop losses.

Suitable plant for providing an adequate quantity of service is available and is being used. The need and demand for service in the rural area is often such as to make it economically attractive to provide it. In providing it then, the problem is that of making the quality agree with the progressive quality of other services available to those in rural areas.

Conclusions

What will be the ultimate transmission objective for subscriber loop losses? Each individual's ear will dictate the answer. A recent cartoon quoted an old man saying, "Us oldsters born before 1900 always holler when we talk on the telephone." The aim today is to design all telephone circuits so that one does not need to "holler" to be heard, and the practice of good circuit design is being carried out in the telephone industry. Progress is indicated by the fact that those who were born in more recent years not only do not "holler," but even whisper when they talk over the telephone and expect to be heard. This quality of service must be provided. Therefore, for the next 10 years −3 to 0 db may be adequate, provided trunk losses are reduced adequately. On the other hand, maybe these loops should be somewhat better, such as −5 to −1.5 db. The question posed will have to wait until time allows the communications engineer sufficient knowledge to examine the "crystal ball" with a more practiced eye. Until that time he must make plans to reduce losses to acceptable values consistent with the exercise of proper caution for an economically sound design. In this respect, it appears that the *E*-type repeater will be in increasing evidence in the long subscriber exchange loop in the immediate future.

References

1. POLYETHYLENE-INSULATED TELEPHONE CABLE, **A. S. Windeler.** *AIEE Transactions*, vol. 73, pt. I, May 1954, pp. 106–11.

2. SIMPLIFIED TRANSMISSION ENGINEERING IN EXCHANGE CABLE PLANT DESIGN, **L. B. Bogan, K. D. Young.** *Ibid.*, Nov. 1954, pp. 498–502.

3. NEGATIVE-IMPEDANCE TELEPHONE REPEATERS—APPLICATION IN THE BELL SYSTEM, **A. F. Rose.** *Ibid.*, pp. 430–35.

4. E-TYPE TELEPHONE REPEATERS—DESCRIPTION, EQUIPMENT, AND TESTING, **J. O. Smethurst.** *Ibid.*, pp. 435–43.

5. THEORY OF E-TYPE REPEATERS, **J. L. Merrill, Jr.** *Ibid.*, pp. 443–47.

Airborne Teletypewriter AN/AGC-1

RICHARD A. MICHALS
NONMEMBER AIEE

Synopsis: The transmission of information by teletypewriter means has proved itself, as shown by the number of civilian and military networks now in operation. The advantage of this method over voice lies in its reduced bandwidth and the production of printed copy. The teletypewriter code also allows specialized data-transmitting techniques to be used; the printed copy eliminates operator stand-by time. It was natural that the Air Force would want to extend its teletypewriter network to include its aircraft. Kleinschmidt Laboratories was awarded the contract to develop such an airborne teletypewriter, known as the *AN/AGC-1*. Basically, the Air Force required a send-receive type of teletypewriter incorporating a recorder-reproducer device for information storage, and a selective call system. This paper presents the results of the developments leading to the design of such a teletypewriter.

THE REQUIREMENT for aircraft installation and operation imposed many problems not normally encountered in teletypewriters. Among these were: 1. Minimum weight and size. 2. Available aircraft space, which required that the set be broken up into major components. 3. Minimum operator time. 4. Aircraft environment. These requirements indicated a new approach to the problem.

The standard printing mechanisms were much too bulky and an isolated unit would be difficult to control. The punched tape method of storing messages had to be changed since it required operator handling. The standard keyboard, receive-transmit mechanism, and control unit would also have to be changed if changes were made in the printer.

It was felt that electronic techniques could be used to replace many of the mechanisms required in a teletypewriter. For example, during receive operation, the changeover of electrical information into mechanical form could be accomplished at the printer selector shaft rather than at the selector magnet. The handling of the information in electrical form, rather than mechanical, reduces the problems.

By using electronic techniques it is possible to build an airborne teletypewriter capable of meeting the requirements of the Air Force. In general, the Air Force required that an airborne teletypewriter provide the following:

1. A page printer capable of operating at 60 WPM (words per minute), alphabetic only.
2. The separation of the set into four major components, known as keyboard, printer, terminal unit, and control panel.
3. All units to be interconnected by means of wires only.
4. A call system for elimination of local traffic.
5. A storage unit to store keyboard information for later transmission at 60 WPM.
6. Operation on 7.5-unit start-stop teletypewriter signals.

Fig. 1 shows the units comprising the *AN/AGC-1*. The purpose of each of the units is as follows:

The control box allows selection of all the operating modes of the teletypewriter. It contains the function switch, power switch, call IN-OUT and rewind indicator light.

The keyboard converts the operator information into teletypewriter form.

The printer converts teletypewriter information into a printed copy.

The terminal unit contains all the electronic hardware for accomplishing the translations.

The first items to be designed were the page printer and the storage unit. The criterion used in their design was to keep the handling of information in mechanical form to a minimum. From these units the rest of the electrical design evolved.

Printer

After an investigation into the various ways of accomplishing the printing, the

---

Paper 58-47, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 15, 1957.

RICHARD A. MICHALS is with Smith-Corona, Inc., Highland Park, Ill.

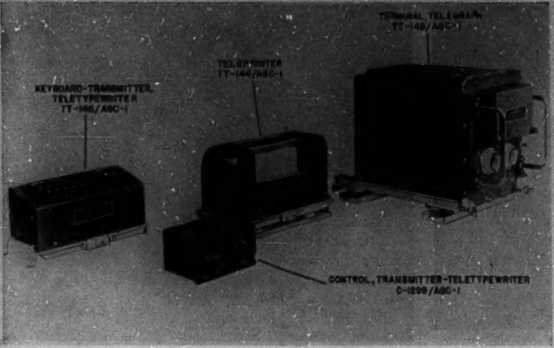


Fig. 1. Airborne teletypewriter set AN/AGC-1

typewheel method was decided upon. In this method, alphabetical letters are engraved on the periphery of a $1^1/_2$-inch-diameter wheel, which is positioned for the correct letter. The wheel is then made to slide on a square shaft; thereby a line of letters is obtained. Pressure-sensitive paper is then held in front of the typewheel, and a print-hammer blow causes the selected letter to be impressed on the paper. Fig. 2 shows the typewheel, square shaft, and print hammer. The paper is held between the print hammer and typewheel.

The positioning of the typewheel is done by a stop pawl engaging a tooth on a ratchet wheel. The ratchet wheel is fastened to the square shaft, along with wipers that act on a printed circuit matrix. The stop pawl is operated whenever the output from the matrix indicates that the desired letter is in position. Fig. 3 shows the stop pawls, the electromagnets for operating them, and the ratchet wheel. The ratchet wheel has 16 teeth; 32 positions are obtained through the use of two stop pawls.

Fig. 4 shows the printed circuit commutator. The commutator works into a diode matrix which receives information from the terminal unit. When a selection is to be made, the square shaft is rotated, which in turn causes the wipers to scan across the printed circuit segments, until a position is reached that corresponds with information fed into the matrix. At that time, the stop pawl is operated and engages a tooth on the stop wheel. This causes the square shaft to stop rotating, also stopping the typewheel which now has the desired letter in front of the print hammer.

Fig. 4 also shows the mechanism for causing the typewheel and print hammer to space or to line return; the paper-to-line feed; and the automatic line return, the line feed, and the printer hammer to operate.

The complete cycle of the printer is:

1. Drive motor shaft is disengaged by positive clutch.
2. Information is supplied to the printer matrix.
3. The printer cycle is started by engaging the positive clutch; this is accomplished by energizing an electromagnet, which also causes the release of the stop pawl, if one is engaged.
4. The engagement of the positive clutch causes the square shaft to rotate.
5. Wipers scan printed circuit segments.
6. Output of matrix indicates when wipers have reached segment corresponding to information supplied to matrix.
7. Matrix output energizes one of the stop pawls.
8. Stop pawl engages tooth on ratchet wheel.
9. Square shaft is stopped; typewheel is positioned according to information fed into matrix.
10. The print hammer is operated.
11. Typewheel spacing and/or line feed and line return mechanisms operate.
12. Positive clutch disengages drive motor shaft. Printer then waits for next cycle.

The square shaft is rotated at 600 rpm

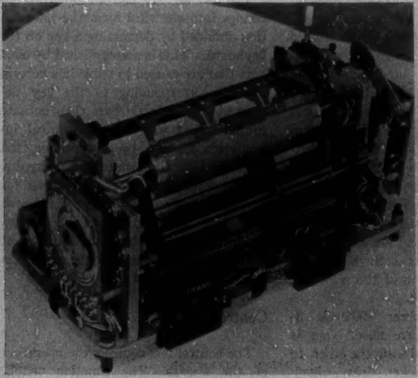


Fig. 2. View of printer with cover removed

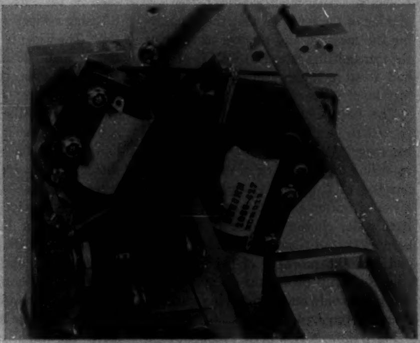


Fig. 3. Stop-pawls, electromagnets, and ratchet wheel

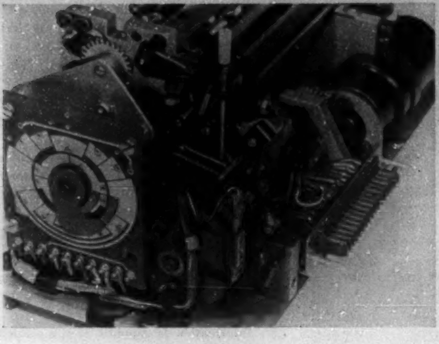


Fig. 4. Printed circuit commutator and mechanisms

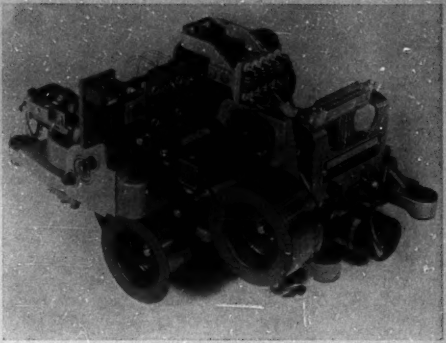


Fig. 5. Magnetic tape recorder

when it is making its selection. The time allowed for the stop pawl to operate is approximately 4 milliseconds. A total of 135 milliseconds is used for printing and the completion of all mechanical functions.

The advantages of such a printer lie in its small size and the ease with which it may be controlled. Its present size is $5^1/_4$ by $5^5/_{16}$ by 11 inches and its weight is 8 pounds. Wires connected to it supply signals for the matrix, mark and space stop-pawl electronmagnets, the positive-clutch electromagnet, and power for the drive motor.

Magnetic Recorder

The storage of teletypewriter information presented several interesting problems. The device had to be capable of storing a message as typed on a keyboard, resetting itself rapidly to the beginning of the message, and then supplying the message any number of times at the transmitting rate of 60 WPM. It also had to store a new message upon completion of a previous message.

The best solution to these problems proved to be the use of a magnetic tape recorder. Magnetic tape allows the recording and playback of messages any number of times. The tape used was made of flutter valve stainless steel, one inch wide, chosen because it could be accurately positioned after long usage, and had the property of being able to wind itself up on a take-up spool. By this means only one tape driving sprocket wheel is required.

Fig. 5 shows a view of the magnetic tape recorder used. The tape may be seen coming off the tape feed spool, passing through the erase coil, passing under the playback-record head and onto the tape sprocket drive wheel, and going to the tape take-up spool. Also shown is the head shuttle arm.

The method used for depositing the bits of information on the tape is to scan the width of the tape. A magnetic recording and playback head is made to travel rapidly across the width of the tape, then to return back to its starting position on the edge of the tape.

This type of operation allows the magnetic head to rest on one edge of the tape during any rewind or waiting-for-information time. Scanning of the magnetic head across the tape is done only when it is desired to record or play back information.

The storage and playback of the five bits (Baudot code) of teletypewriter information requires the separating of each of the bits. This is accomplished by using a wiper which scans a commutator. The wiper is mounted to move with the magnetic head. As the head sweeps across the tape, the wiper moves across the commutator segments, thereby separating each of the bits. To record, voltage information is supplied to the commutator and the recording head is connected to the wiper. For playback, the output of the head amplifier is connected to the wiper, and the commutator segments to a buffer storage. Fig. 6 shows a block diagram of the recording and playback circuits.

Two 1/150-horsepower 400-cycle 3-phase drive motors are used. One is used to cause the head scan, the other, to drive the tape sprocket drive wheel. This latter motor is reversed when it is desired to rewind the tape. All the motors operate through friction clutches. The cycling of the recorder is controlled by an electromagnetic latching mechanism. Releasing the latch allows the head to scan across the tape and to return, and causes the tape sprocket wheel to space the tape. At 60 WPM, 294 inches of tape are required for a storage time of 10 minutes.

This design has resulted in a rugged storage unit for recording and playback of binary information. Because it is mounted with electronic tubes it had to be capable of operating at a temperature of 95 degrees centigrade. It is now being used in other systems requiring these capabilities.

Keyboard

The purpose of the keyboard is to change into electrical form the information obtained by depressing a key on the keyboard. This is accomplished by code bars that are caused to move in a certain manner corresponding to the key depressed. These code bars then actuate contacts that will supply electrical information for use by the printer, the recorder, and the transmitting circuits. The code bars are counterbalanced for operation under conditions of extreme vibration. Included in the keyboard is a locking mechanism that is used to prevent the operator from exceeding the maximum transmitting speed of 60 WPM.

Control Box

The control box controls the operation of the system. The function swtch. determines the mode of operation, i.e. transmit-store, transmit-keyboard, re-

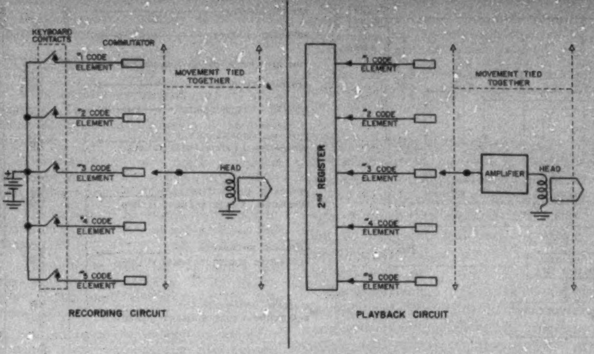


Fig. 6. Block diagram of recording and playback circuit

ceive, rewind, and store. The call switch allows the operator to bypass the call system and allows him to print all messages that are received. The power switch controls all the voltages going to the system. The rewind light indicates when the tape in the storage unit is in its fully rewound position.

The modes of operation of the control box are:

Transmit-store: The teletypewriter set will automatically transmit any message previously recorded on the recorder. At the end of message the set will automatically go into "receive" condition. A page copy of the message is also printed.

Transmit-keyboard: Transmission of messages is made directly from the keyboard and is monitored by the printer.

Receive: With the call switch in the out position, any messages received from the radio equipment will be printed. With the call switch in the in position, only messages preceded with a call will be printed.

Rewind: The magnetic tape recorder is operated to rewind the tape to its starting position. When the tape is in this position, the rewind light will light.

Store: A message will be recorded on the magnet recorder by operating the keyboard. The recorded message will not be transmitted; the printer will monitor the recorded message. If a message containing a call is received, the keyboard will automatically lock up, and the printer will print the received message. After the received message is completed, the keyboard will unlock and the recording may be resumed.

Terminal Unit

Almost all of the electronic circuits for operating the teletypewriter are contained inside the terminal unit. The magnetic storage unit is also housed in the terminal unit. The dimensions of the outer case are $10^3/_4$ by $10^1/_8$ by 15 inches, and the weight of the unit is 29 pounds.

Fig. 7 shows a view of the terminal unit removed from its case. The placement of the storage unit and the three subchassis may be seen. On the front panel is the call patch box and the connectors used in supplying signals to and from the terminal unit. Under the subchassis location are the relays and tantalum capacitors (used for filtering the power supply). Under the storage unit location are the 400-cycle power transformer, the silicon rectifiers and voltage-regulator tubes, the reed oscillator, and the polar relay.

All the subminiature tubes are mounted on the three aluminum subchassis. Resistors and capacitors are wired into the circuit by means of teflon-insulated stand-offs. For servicing, these subchassis may be removed from the terminal unit and repaired on the bench.

The circuits contained inside the terminal unit are used to accomplish the functions normally done mechanically. The most important of these is the generation of an accurate sequential signal. This is used for generating the sequential signal for transmitting, and for sensing the incoming sequential signal code elements. This is accomplished by an 8-tube linear ring counter driven by a time base.

Fig. 8 is a block diagram of the transmit-keyboard circuit. It consists of the following:

1. A reed-controlled oscillator operating at 272.7-cps (cycles per second).
2. First divider which is driven by the 272.7-cps oscillator and divides to 90.9 cps.
3. Second divider, driven by the first divider, dividing to 45.45 cps except during the time the no. 7 ring output is energized. At this time the output is 90.9 cps.
4. Ring counter distributor.
5. Gating circuit controlled by keyboard for forming sequential signal.
6. Polar relay for output.

A cycle of operation is as follows:

1. Ring counter is normally at its rest

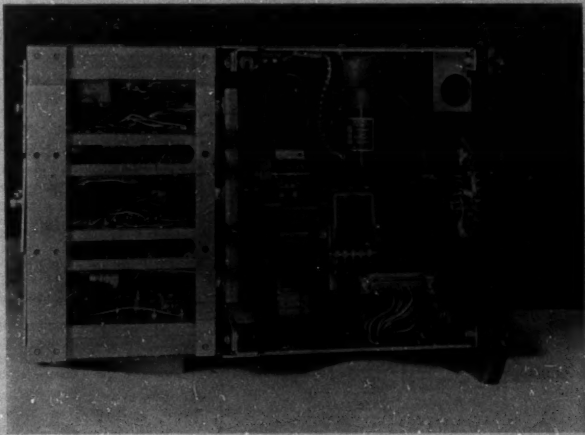
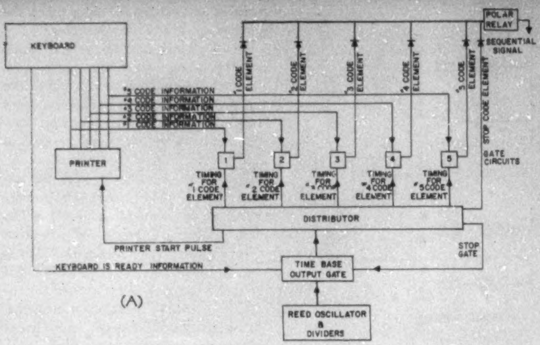


Fig. 7. Top view of terminal unit removed from case


KEYBOARD
POLAR RELAY
SEQUENTIAL SIGNAL
#1 CODE ELEMENT
#2 CODE ELEMENT
#3 CODE ELEMENT
#4 CODE ELEMENT
#5 CODE ELEMENT
STOP CODE ELEMENT
#5 CODE INFORMATION
#4 CODE INFORMATION
#3 CODE INFORMATION
#2 CODE INFORMATION
#1 CODE INFORMATION
PRINTER
1
2
3
4
5
GATE CIRCUITS
TIMING FOR #1 CODE ELEMENT
TIMING FOR #2 CODE ELEMENT
TIMING FOR #3 CODE ELEMENT
TIMING FOR #4 CODE ELEMENT
TIMING FOR #5 CODE ELEMENT
DISTRIBUTOR
PRINTER START PULSE
STOP GATE
KEYBOARD IS READY INFORMATION
TIME BASE OUTPUT GATE
REED OSCILLATOR & DIVIDERS
(A)

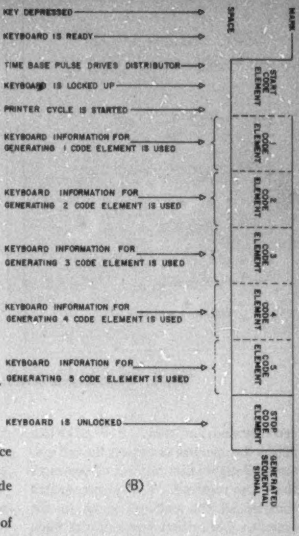

KEY DEPRESSED
KEYBOARD IS READY
TIME BASE PULSE DRIVES DISTRIBUTOR
KEYBOARD IS LOCKED UP
PRINTER CYCLE IS STARTED
KEYBOARD INFORMATION FOR GENERATING 1 CODE ELEMENT IS USED
KEYBOARD INFORMATION FOR GENERATING 2 CODE ELEMENT IS USED
KEYBOARD INFORMATION FOR GENERATING 3 CODE ELEMENT IS USED
KEYBOARD INFORMATION FOR GENERATING 4 CODE ELEMENT IS USED
KEYBOARD INFORATION FOR GENERATING 5 CODE ELEMENT IS USED
KEYBOARD IS UNLOCKED
SPACE
MARK
START CODE ELEMENT
1 CODE ELEMENT
2 CODE ELEMENT
3 CODE ELEMENT
4 CODE ELEMENT
5 CODE ELEMENT
STOP CODE ELEMENT
GENERATED SEQUENTIAL SIGNAL
(B)

Fig. 8. A—Block diagram for transmit-keyboard circuits. B—Timing chart for transmit-keyboard operation

position. Action of the time base pulses are blocked by the start gate tube.

2. Depressing a key on the keyboard allows the ring counter to make one cycle, and sets up the gating circuits used in forming the sequential signal to be transmitted.

3. The various outputs from the ring counter are used in the keyboard gating circuits for driving the polar relay, and for starting the printer cycle.

The timing chart shows the sequence of operation.

Fig. 9 is a block diagram of the receive circuit. It consists of the following:

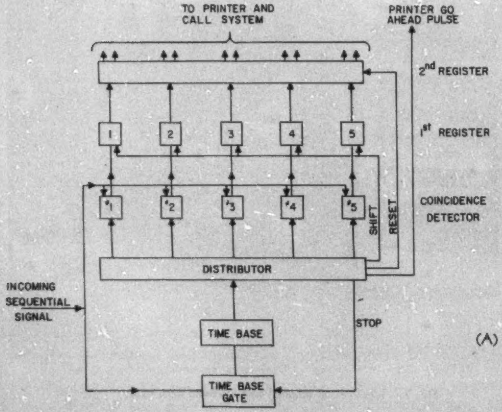
1. Multivibrator receive time base for driving the distributor.

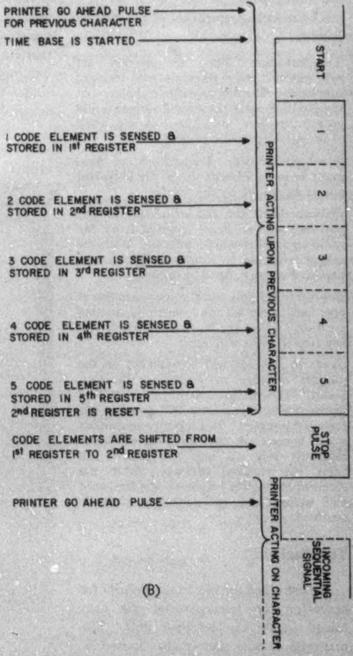
2. Ring counter distributor for obtaining timed sensing pulses used in coincidence detectors.

3. Coincidence amplifier for sensing code elements.

4. First register: temporary storage of code elements.

Fig. 9. A—Block diagram for receive circuit. B—Timing chart for receive operation


TO PRINTER AND CALL SYSTEM
PRINTER GO AHEAD PULSE
2nd REGISTER
1st REGISTER
1
2
3
4
5
#1
#2
#3
#4
#5
SHIFT
RESET
COINCIDENCE DETECTOR
DISTRIBUTOR
INCOMING SEQUENTIAL SIGNAL
TIME BASE
STOP
TIME BASE GATE
(A)


PRINTER GO AHEAD PULSE FOR PREVIOUS CHARACTER
TIME BASE IS STARTED
1 CODE ELEMENT IS SENSED & STORED IN 1st REGISTER
2 CODE ELEMENT IS SENSED & STORED IN 2nd REGISTER
3 CODE ELEMENT IS SENSED & STORED IN 3rd REGISTER
4 CODE ELEMENT IS SENSED & STORED IN 4th REGISTER
5 CODE ELEMENT IS SENSED & STORED IN 5th REGISTER
2nd REGISTER IS RESET
CODE ELEMENTS ARE SHIFTED FROM 1st REGISTER TO 2nd REGISTER
PRINTER GO AHEAD PULSE
PRINTER ACTING UPON PREVIOUS CHARACTER
PRINTER ACTING ON CHARACTER
START
1
2
3
4
5
STOP PULSE
INCOMING SEQUENTIAL SIGNAL
(B)

5. Second register: storage of code elements for use by printer.

A cycle of operation is as follows:

1. Ring counter is normally at its rest position. The time base for driving the distributor is blocked by action of the start gate tube.

2. The transition from mark to space caused by a start pulse being received is used to start the time base.

3. The time base drives the distributor in synchronism with the incoming sequential signal.

4. The outputs from the distributor are used in the coincidence amplifier to determine the mark condition of the five incoming code elements.

5. Output from the coincidence amplifier is stored sequentially in the first register.

6. After the five code elements have been stored in the first register, they are transferred simultaneously into the second register.

7. The second register supplies the printer matrix with information which it then acts upon. The call system also receives its information from the second register.

8. The distributor comes back to its rest position, thereby stopping the time base. The cycle is completed.

The formation of the sequential signal for transmit-store operation is much the same as transmit-keyboard. Information from the storage unit is fed into the second register, which then supplies the information the keyboard had been supplying. It is then handled in the same manner as the keyboard information.

The selective call system makes use of a straight counting circuit. If the counter completes its count before any resetting pulses occur, it activates the operation of the printer. These resetting pulses are caused by space signals which are always present in a message. The letters that the call counter will respond to are selected by means of the matrix box which is plugged into tube sockets wired for the desired letters.

Subminiature thyratrons are used to energize the recorder and printer-selector electromagnets. A beam power subminiature tube is used to energize the printer clutch electromagnet. This tube also extinguishes any printer selector thyratrons that might have been fired.

A total of 45 tubes are used in the system. Thirty-nine are subminiature tubes and six, miniature tubes. The system uses 120 watts of 400-cycle power and 60 watts of d-c power.

Conclusions

The use of teletypewriter equipment aboard aircraft will increase the reliability of long-range communications. It will also greatly reduce the operator's time for transmitting and receiving messages. These features are being required more and more in the newer aircraft.

Automatic Telegraph Switching System Plan 55-A

G. S. VERNAM
MEMBER AIEE

THE 200,000-mile private-wire network of the United States Air Force includes five Plan 51 switching centers in the United States, leased from The Western Union Telegraph Company. These centers are interconnected by trunk lines and each center has lines radiating to tributary stations in its area. Gateway centers near the Atlantic and Pacific coasts extend the network over radio channels to numerous Air Force installations overseas. This network has been in operation since February 1951. Satisfactory experience in operating this network, with its push-button-operated Plan 51 switching centers, has led the Air Force to request Western Union to engineer a fully automatic switching system for world-wide use. It is expected that automatic switching will increase the speed and efficiency of message handling and result in considerable reduction in operating labor. It will also provide for additional circuit capacity.

This new switching system, Fig. 1, known as Plan 55-A, has been developed for the Air Force for handling military traffic. It is expected that Plan 55-A automatic switching centers will be installed to replace the present Plan 51 centers in the United States and for overseas use, as soon as equipment is available.

To meet the needs of the U. S. Air Force for rapid and secure transmission of very large volumes of military traffic over a world-wide network, a switching system should meet requirements as follows:

1. It should provide for rapid automatic switching of messages through any number of switching centers without operating attention.

2. Where defects in a message or trouble conditions prevent automatic switching, means should be available for immediate manual (push-button) switching of messages to avoid delay.

3. The switching system must be compatible in operation with the equipment used by the other military services, so that messages may be relayed from one service to another without manual retyping. Obviously, the switching system must also be compatible with any other message-relaying system in use by the Air Force.

4. The switching equipment must be capable of either automatically or manually switching multiple-address messages.

5. The equipment should be designed on the "building block" principle, with the minimum number of different types of cabinets, each complete in itself, with plug-in connections to the other cabinets arranged for quick installation, or quick removal and reinstallation at another location of all or part of a switching center as may be required. No equipment that is essential for switching should be used in common by the entire switching center.

6. Fast handling of all traffic is essential, with preference in switching being given to messages of high precedence.

The Plan 55-A switching system, developed by Western Union to meet these requirements, consists of groups of equipment units installed in switching centers interconnected over a network of lines or radio channels to form an operating system for automatically relaying telegraph messages. A typical Plan 55-A network, such as that shown in Fig. 2, has a number of reperforator switching centers interconnected by trunk lines, with lines radiating from each center to tributary stations in the area served by that center. Messages originated at any tributary station are relayed automatically through one or more switching centers to their destina-

---

Paper 58-45, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing November 15, 1957.

G. S. VERNAM is with The Western Union Telegraph Company, New York, N. Y.

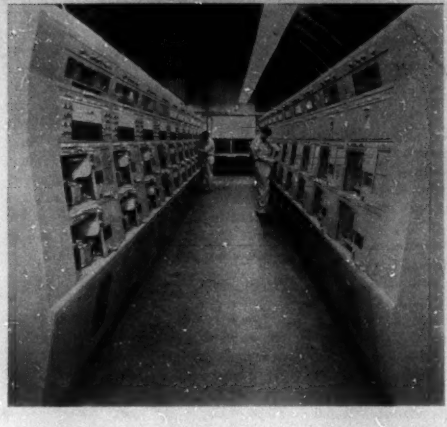


Fig. 3. Sequence number indicator

Fig. 1 (left). Plan 55-A prototype set up for test

tions at other tributary stations. This system may be operated over a nationwide network of trunk lines or radio channels and may be expanded beyond national and continental boundaries. A Plan 55-A switching center can exchange traffic with switching centers using other types of switching equipment, including Plan 51 reperforator switching or semiautomatic, "torn-tape," relaying equipment.

In Plan 55-A switching centers, messages are received and retransmitted automatically in the form of perforated tape. Routing is controlled by "routing indicators" (normally groups of six letters) recorded in the tape as part of the heading of each message. The form of message

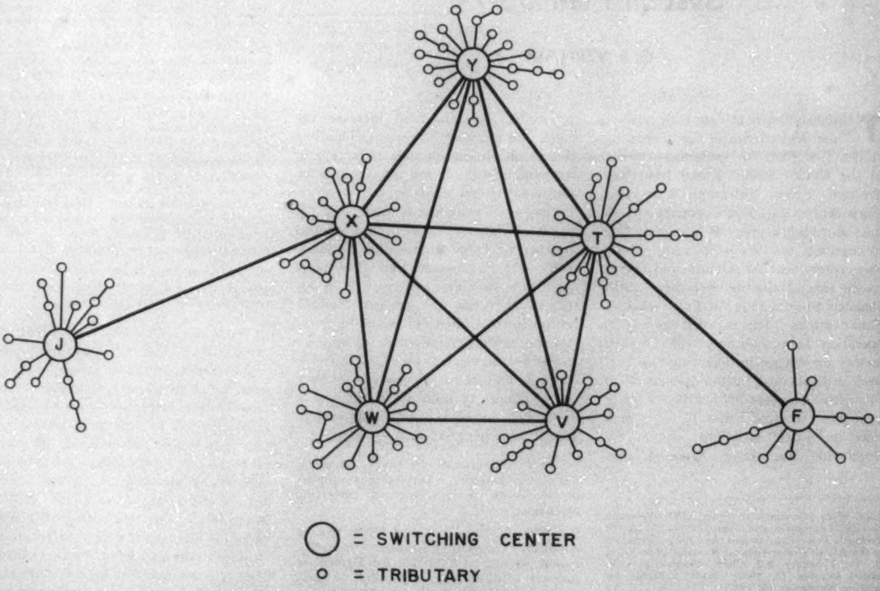


Fig. 2. Hypothetical network

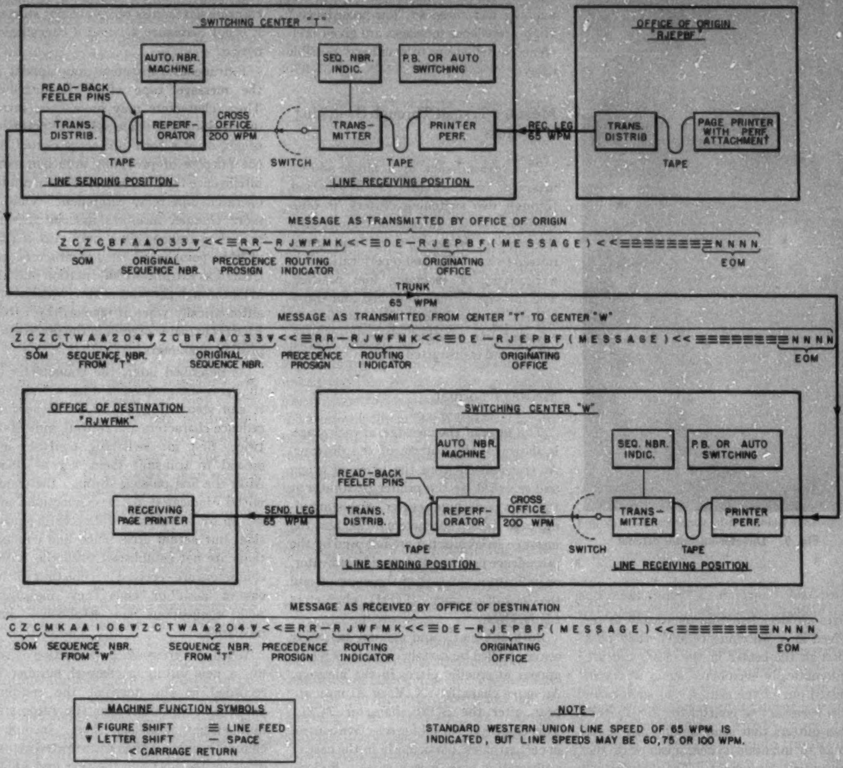

SWITCHING CENTER "T"
AUTO. NBR. MACHINE
SEQ. NBR. INDIC.
P.B. OR AUTO SWITCHING
READ-BACK FEELER PINS
TRANS. DISTRIB.
REPERF-ORATOR
CROSS OFFICE 200 WPM
TRANS-MITTER
PRINTER PERF.
TAPE
SWITCH
TAPE
LINE SENDING POSITION
LINE RECEIVING POSITION
REC. LEG 65 WPM
OFFICE OF ORIGIN "RJEPBF"
TRANS. DISTRIB.
PAGE PRINTER WITH PERF. ATTACHMENT
TAPE
MESSAGE AS TRANSMITTED BY OFFICE OF ORIGIN
Z C Z C B F A ▲ 0 3 3 ▼ < < ≡ R R — R J W F M K < < ≡ D E — R J E P B F (M E S S A G E) < < ≡≡≡≡≡≡≡≡ N N N N
SOM
ORIGINAL SEQUENCE NBR.
PRECEDENCE PROSIGN
ROUTING INDICATOR
ORIGINATING OFFICE
EOM
TRUNK 65 WPM
MESSAGE AS TRANSMITTED FROM CENTER "T" TO CENTER "W"
Z C Z C T W A ▲ 2 0 4 ▼ Z C B F A ▲ 0 3 3 ▼ < < ≡ R R — R J W F M K < < ≡ D E — R J E P B F (M E S S A G E) < < ≡≡≡≡≡≡≡≡ N N N N
SOM
SEQUENCE NBR. FROM "T"
ORIGINAL SEQUENCE NBR.
PRECEDENCE PROSIGN
ROUTING INDICATOR
ORIGINATING OFFICE
EOM
OFFICE OF DESTINATION "RJWFMK"
RECEIVING PAGE PRINTER
SEND. LEG 65 WPM
SWITCHING CENTER "W"
AUTO NBR. MACHINE
SEQ. NBR. INDIC.
P.B. OR AUTO SWITCHING
READ-BACK FEELER PINS
TRANS. DISTRIB.
REPERF-ORATOR
CROSS OFFICE 200 WPM
TRANS-MITTER
PRINTER PERF.
TAPE
SWITCH
TAPE
LINE SENDING POSITION
LINE RECEIVING POSITION
MESSAGE AS RECEIVED BY OFFICE OF DESTINATION
C Z C M K A ▲ 1 0 6 ▼ Z C T W A ▲ 2 0 4 ▼ < < ≡ R R — R J W F M K < < ≡ D E — R J E P B F (M E S S A G E) < < ≡≡≡≡≡≡≡≡ N N N N
SOM
SEQUENCE NBR. FROM "W"
SEQUENCE NBR. FROM "T"
PRECEDENCE PROSIGN
ROUTING INDICATOR
ORIGINATING OFFICE
EOM
MACHINE FUNCTION SYMBOLS
▲ FIGURE SHIFT
≡ LINE FEED
▼ LETTER SHIFT
— SPACE
< CARRIAGE RETURN
NOTE
STANDARD WESTERN UNION LINE SPEED OF 65 WPM IS INDICATED, BUT LINE SPEEDS MAY BE 60, 75 OR 100 WPM.

Fig. 4. Switching single-address message through two switching centers

conforms to standards agreed upon among the Armed Services, and the switching equipment has been designed to be compatible in operation with the equipment which is being used by the other military services.

Messages may be switched automatically through several switching centers without requiring the attention of operators, except at the points or origin and destination. However, each receiving position at a switching center may be converted to manual (push-button) switching, at any time, by operating an "auto-manual" switch. It is expected that automatic switching will be the normal operating method for use in all cases where the quality of reception at the switching center and of tape preparation at the point of origin permits. However, in case of failure to switch automatically, because of incorrect routing indicators, or other errors in the received message tape, a supervisory signal is operated to call in a switching attendant so that the message may be manually switched without delay. It is also contemplated that small centers might be set up quickly in an emergency, with the automatic switching equipment omitted to save time in installation.

Message Security

Numerous safeguards and alarms are provided throughout the system to prevent errors, avoid delays, and guard against lost messages. Each message is numbered in sequence as transmitted to an outgoing line or channel and these channel numbers are automatically checked as the messages are relayed. For this purpose, automatic message-numbering machines are provided to transmit the office call, channel designation, and serial number ahead of each message. For each incoming circuit a number comparator device, known as a sequence-number indicator (SNI), is provided to check these numbers automatically. The SNI is similar in design to the automatic numbering machine and is shown in Fig. 3.

It is contemplated that some messages will be transmitted in code or cipher. For this purpose either on-line or off-line automatic ciphering equipment can be used. This cryptographic equipment is not part of the Plan 55-A switching equipment, and is not discussed in detail. With on-line cryptographic equipment, mes-

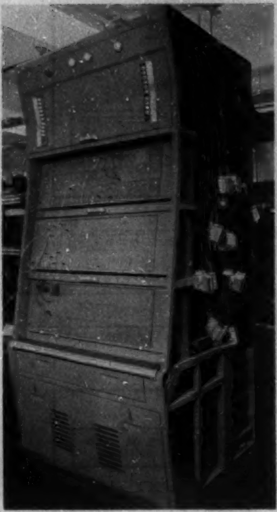


Fig. 5. Director-translator cabinet

sages are deciphered automatically as received in a switching center; they pass through the center in the clear, and are automatically enciphered again as transmitted out of the center. In such cases it is necessary to control the distributor-transmitters that send to outgoing lines, so as to maintain synchronism with the encrypting equipment. The distributor-transmitters are of a type permitting this control. With off-line cryptographic equipment, messages are enciphered at the point of origin, transmitted through the switching centers in cipher, and deciphered at the point of destination. Such messages must have clear headings for routing purposes and must avoid the chance occurrence of the message-termination character combination (normally LF NNNN) in the ciphered message text.

Message Precedence

Messages are transmitted from the office of origin in order of precedence and are delivered at their destinations in the same order. There are six precedence classifications designated in the message by repeated characters called "precedence prosigns." For switching purposes the first three precedence classifications are normally considered as "high-precedence" and the last three as "low-precedence." High-precedence messages are given preference in switching to avoid any possible delays.

Message Transit Through Switching Centers

The transmission of a single-address message from origin to destination, through two switching centers, is illustrated in Fig. 4. At each switching center the message is reperforated and transmitted twice. The first reperforation is at a receiving or incoming line position. The message is then switched and transmitted "across office" to a sending or outgoing line position where it is reperforated and transmitted again.

Message Format

The form of the message, at each stage, is shown at the bottom of the drawing. As transmitted from the office of origin and recorded by the printer-perforator at the first switching center, the start-of-message (SOM) characters ZCZC and the message serial number are followed by the precedence prosigns and routing indicator, or indicators, the text of the message and the end-of-message (EOM) characters LF NNNN.

Both automatic and manual switching are controlled by certain characters which appear at specific places in the message. An extra character (X, Y, or Z) may appear after the SOM character ZCZC. This character is used for selective control in certain cases, particularly in the case of multistation lines or way wires. Following the end of the preceding message, the first significant characters of the next message are ZCZC. Intervening characters which may have been inserted by the printer-perforator for feed-out, or by the station of origin, are treated as extraneous characters that are stepped through the transmitter and discarded.

The incoming message channel number consists of circuit-identifying letters and a 3-digit serial number. These characters are checked by the SNI, before the message is switched. Checking each received message serial number, as it appears in the tape, prevents loss of messages, facilitates identification of messages for reruns and also serves as a periodic transmission check on the line and on the operation of the transmitter and printer-perforator. In case of a wrong number comparison, transmission is stopped immediately and supervisory alarm signals are actuated. As this occurs before the message has been switched, and as transmission across office has not started, it is not necessary to send a cancellation notice.

Extraneous characters may appear in the message tape after the number. These characters may have been introduced into the tape at the point of origin, or at a torn-tape relay center, for the purpose of providing sufficient non-intelligence tape for insertion in manually operated tape relay equipment. One or more channel numbers associated with preceding transmissions may also appear at this point. All such characters are treated as extraneous information and are idled through the transmitter, which stops automatically when it reaches the LINE FEED character preceding the precedence prosign characters.

As described later, the transmitter at the receiving position is designed so that it can step through and read the precedence characters and routing indicators twice, first for switching control, and second to transmit them across office. After the first passage through the transmitter, indicated route connections are set up by switches in the receiving position, but actual cross-office line connections are not established until the EOM characters are received, except that in the case of "flash" or "emergency" messages, such connections are established immediately.

When the cross-office connection is set up, a new outgoing channel number is recorded in the tape at the sending position, transmitted from the automatic numbering machine. The incoming channel number is then transmitted across office by the SNI and the text and EOM characters are sent by the transmitter. After the EOM characters the cross-office circuit is disconnected.

From the sending position the message may be sent to a tributary of the same switching center, but in most cases the message will be switched through at least two centers. The message, as received by the printer-perforator at the second center, is the same as when it was received at the first center except that it now has two serial numbers, an outgoing channel number, and an incoming channel number. The operations involved in transmitting the message through the second center are the same as those just described for the first center, except that when the first number is checked by the SNI, the second number steps through the transmitter and is discarded. A new outgoing channel number is added at the head of the message by the numbering machine and the former outgoing channel number becomes the incoming channel number at the second center. Only two channel numbers are carried along with the message as it is relayed from center to center, thus providing adequate identification while conserving line time.

Multiple-Address Messages

Multiple-address messages are recognized as such by the presence of more than one routing indicator. Each routing indicator is preceded by a space character and the last routing indicator is followed by a carriage return character. All routing indicators are printed in one line and the number of routing indicators in a message is limited to nine. Multiple-address messages are automatically processed so as to transmit an individual copy of the message to each ultimate destination. In this process the routing indicators are segregated so that each final copy of the message contains only one routing indicator, corresponding to its destination. At a switching center several of the routing indicators in a message may require the same circuit outlet. For example, two or more routing indicators may be for tributaries of another switching center, or for more distant centers to be reached through the same intermediate switching center. Such messages are transmitted to the intermediate center containing only the routing indicators for which that center is responsible.

Any receiving position can switch a multiple-address message directly to not more than four outlets. As relatively few messages require switching to more than four circuit outlets, most messages may be switched directly from the receiving positions. All preliminary operations for a multiple-address message are the same as for a single-address message up to the point where the first routing indicator has been processed. The space character following the first routing indicator is identified and the transmitter is restarted to process the second routing indicator. This process continues until a carriage return character, after the last routing indicator, identifies the end of the routing line.

In relatively rare cases, a message may need to be switched to more than four circuit outlets. In such cases, all routing indicators requiring an outlet different from the first three, are sent with a copy of the message to an extra operating position assigned for this purpose. The messages as recorded at this extra position contain at least three fewer routing indicators than appeared in the original message. In very rare cases a message with eight or nine routing indicators may need to be transmitted to an extra position a second time.

Switching Center Components

The two principal items of equipment required in a Plan 55-A switching center are the incoming and outgoing line consoles or cabinets. Each cabinet houses two operating positions. The incoming line cabinet has a control cabinet installed behind it to provide additional mounting space for equipment. These cabinets are interconnected by cables terminated in multiconductor plugs and sockets. To meet military requirements, all cabinets of each type are alike and require no essential common equipment to set up an operable center for manual switching. The incoming and outgoing cabinets are linked together by cross-office switching circuits over which messages are transmitted electronically at 200 WPM (words per minute). In Fig. 1, the incoming or receiving cabinets are at the left and the outgoing or sending cabinets at the right side of the aisle.

These cabinets contain all the equipment needed for reperforating, transmitting, and push-button switching. The building-block principle thus obtained enables a center to be expanded or contracted as circumstances dictate with the minimum of disturbance to traffic handling. As initially equipped, each incoming line position can switch messages to a maximum of 200 circuits and 100 destinations, by providing a suitable number of outgoing line consoles. By adding switches, the equipment can be expanded to 400 circuits and 200 destinations.

A third type of equipment, a director-translator cabinet, shown in Fig. 5, is required for automatic switching. The cabinet houses two directors and a translator. It is similar in design to the receiving cabinets and is normally installed at the end of a receiving row. The director receives routing information from the message tape and it controls switching operations at the start of each message. The translator receives routing indicators from the director and translates them into switch settings. As these operations take only a few seconds, a relatively few directors and translators are required. It is proposed to install them in the ratio of one director cabinet for every 25 incoming line cabinets.

The cabinets of all three types are 84 inches high, 38 inches wide and 33 inches deep. Associated with each receiving or incoming line cabinet is a receiving control cabinet, shown at the left in Fig. 6, which contains rotary switches and other control equipment. The control cabinet is installed directly behind its associated receiving cabinet, with a maintenance aisle between. To conserve floor space, the control cabinets may be installed against a wall or another row of cabinets. A similar arrangement is used for the director cabinet, with two directors housed in the rear cabinet and a translator and associated connecting relays in the front cabinet.

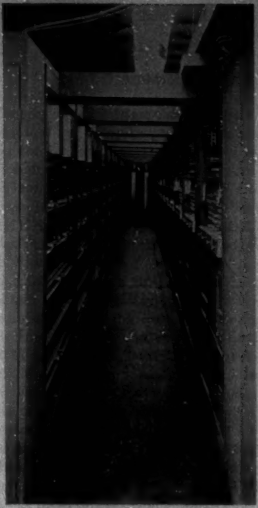


Fig. 6. Maintenance aisle behind receiving cabinets

In addition to the three types of cabinets just described, certain common equipment, is normally provided at a "traffic control center." Although a switching center may be operated without this common equipment, it is desirable to provide it to facilitate traffic handling and supervision. This equipment includes a connection indicator board, a traffic-routing board, a close-out indicator board, and supervisory printer sets, to be described later.

Receiving Positions

Each incoming line cabinet contains equipment for two receiving positions. Either position may be used to terminate an incoming trunk or single-station tributary circuit. The upper position may be used also to terminate a multistation circuit or to terminate a trunk having separate high- and low-precedence number sequences.

Printer-Perforators

Each receiving position is equipped with type 28 printer-perforator for reproducing incoming messages in printed-perforated form; see Fig. 7. Associated circuits designed to detect and count bell signals (upper case S), the SOM signal ZCZC and the EOM signal LF NNNN. The bell signals are used to detect "flash" and "emergency" messages. The EOM signal controls the message-waiting indicator (MWI), adding one on its dial for each message received. The transmitter circuit subtracts one for each complete message transmitted, so that the MWI dial indicates the number of complete messages on hand that have not been completely transmitted. The MWI also controls tape metering and is so arranged that blank tape will be fed out if incoming reception stops after the EOM, if the MWI indicates not more than one message on hand. The SOM and EOM signals are received alternately, as they occur at the start and end of each message. If the EOM signal is missing or defective, as received, there will be two successive SOM signals. This disables the MWI ADD circuit and operates a no-EOM alarm, when the MWI pointer restores to zero. This action blocks cross-office transmission, thus preventing a message from following through to the same destination after a preceding message with a defective EOM combination. The operating attendant restores service by restarting the transmitter, stopping it again at the end of the message and then operating an XMIT EOM switch to initiate automatic transmission of a correct EOM signal.

Loop-Gate Transmitters

Each receiving position has a loop-gate transmitter, shown in Fig. 8, for reading characters in the message tape and transmitting them at 200 WPM to the automatic switching director or across office to a reperforator at a sending position. Associated with the transmitter is a SNI for checking incoming message numbers, as recorded in the tape, and retransmitting them across office at the proper time. The transmitter also has associated electronic means for detecting certain characters in the tape as they pass through the transmitter. A motor-driven tape winder is provided to store message tape after transmission. An electronic transmitter is also provided to transmit messages and to transmit and receive control signals across office.

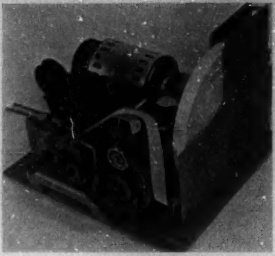


Fig. 7. Type 28 printer-perforator for incoming messages

The loop gate is a slidable latch which holds down the tape and permits sending certain portions of the message twice. With the loop gate in its right-hand position the transmitter functions, in the same way as any conventional magnet-stepped transmitter, to idle blank tape through, past the sensing pins. The SOM characters ZCZC start the SNI which checks the call letters and message number in the tape. If the message number is incorrect, the transmitter stops and a wrong-comparison signal lights to call an attendant who will take necessary corrective action. If the message number is correct, the transmitter steps the tape ahead until it detects a line feed (LF) character, thus stepping through and deleting any extra numbers or other extraneous characters that may appear in the tape at this point. Upon detecting the line feed, the loop gate on the transmitter slides to the left, holding the tape so that it cannot pass completely through the transmitter, but can step past the sensing pins forming a loop above the transmitter. Characters fed into this loop may be transmitted a second time.

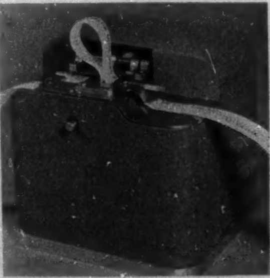


Fig. 8. Loop-gate transmitter for sending messages across office

Automatic Switching

For automatic switching, with the loop gate at the left, a connection is obtained to a director and the precedence prosign characters and the routing indicator or indicators are transmitted to the director, which returns switching information to the receiving position, for setting up the cross-office connection to the desired destination. The loop gate then returns to its right-hand position, placing the original character (the first character after the line feed) over the sensing pins. The transmitter is now ready to transmit the message across office at the proper time.

Manual Switching

For manual switching, when the loop gate slides to the left, a message-waiting (MSG WTG) lamp lights to call the attendant, who looks at the precedence prosign characters on the tape and depresses a high- or low-precedence push-button, which associates the transmitter with the push-button panel, advances the prosign characters through the transmitter, extinguishes the MSG WTG lamp and lights a "switching control" lamp. The attendant now depresses a panel push-button for each routing indicator and then a START push-button to complete the manual switching operations.

Electronic Cross-Office Transmission

Electronic methods are used for transmitting signals across office over a single conductor. The electronic transmitting and receiving equipments are synchronized to the 60-cycle power supply. The message characters are transmitted at 200 WPM (20 characters per second), which is equivalent to one character for every third cycle of the 60-cycle power supply. Certain control pulses are transmitted over the same cross-office line conductor, in either direction, at the rate of 60 pulses per second. As compared to other switching systems, the use of a single conductor, in this system, greatly reduces the number of cable conductors, multiconductor plugs, sockets, and switches required. For example, eight levels of a single 25-point rotary switch serve to select any one of 200 sending positions.

The electronic units consist essentially of a pulse generator, common to the two operating positions in each cabinet, an electronic transmitter for each receiving position and an electronic receiver for each sending position. These units generate, transmit, and detect the message and control pulses that are sent across office and, in addition, step the tape transmitter and numbering machine, detect and indicate trouble conditions, read certain message characters as they are being transmitted at high speed, test the idle or busy condition of selected routes, and control the operation of other equipment, such as the SNIs, the MSG WTG indicators, and the automatic switching directors. The pulse generator and electronic transmitters are shown at the top of Fig. 9.

Sending Position Cabinets

Each outgoing line cabinet contains equipment for two sending positions. The upper position may be used to terminate any type of outgoing line, including a multipoint circuit (way wire). The lower position can terminate any line except a multipoint circuit. Either position may be used as part of a multichannel group.

Sending Positions

Each sending position is equipped with a Type 28 multimagnet reperforator, shown in Fig. 10, for reproducing messages received over cross-office circuits in perforated (not printed) tape form. This reperforator is associated with an electronic receiver which operates on signals received across office at 200 WPM as sent from an electronic transmitter associated with the loop-gate transmitter at a receiving position. The reperforator also receives signals from an automatic message numbering machine at the same sending position. In the case of multipoint circuits, terminated at upper sending positions, there may be two or three numbering machines (one for each way station), any one of which can send into the reperforator.

When a cross-office connection is established, the numbering machine automatically sends, into the reperforator, the SOM characters ZCZC, a multistation selector code character (X, Y, or Z) if required, the station and channel designation letters, and the message number. This number is followed by the incoming channel number transmitted from the SNI at the receiving position. The latter number is always preceded by the characters ZC and followed by CR CR LF. The message is then transmitted across office by the loop-gate transmitter, first the precedence prosign characters and routing indicators from the tape loop and then the remainder of the message, including the EOM characters LF NNNN. At this point, the loop-gate transmitter stops, the numbering machine sends two blanks into the reperforator, and the cross-office circuit is released.

Character Checking

The reperforator is equipped with read-back contacts operated by an extra set of sensing pins located five characters to the left of the punch pins. These pins read certain character groups in each message to verify the correct operation of all equipment involved in cross-office transmission. The SOM characters ZCZC from the numbering machine are the first characters checked. The appearance of this combination indicates that the local equipment of the sending position is operating properly, including the numbering machine and the reperforator. The ZC characters ahead of the incoming channel number verify the operation of the SNI and cross-office transmission. At the end of the message, when two blanks are transmitted by the numbering machines after the EOM characters, the line feed and first N are fed over the read-back pins and checked, thus verifying that the reperforator is still operating properly at the end of the message. The cross-office connection is then released.

The ZC characters must check to start cross-office transmission of the message. Failure to check prevents transmission, locks out the sending position, and operates an alarm signal at the sending position and a reswitch signal at the receiving position. A number-cancellation notice is then transmitted automatically from the receiving position. The sending position remains out of service until manually restored after the trouble is cleared. At the receiving position an attendant sets the tape back to the start of the message and releases it to be reswitched. It will be automatically switched to the same destination through another sending position, if one is available, or to an intercept position for storage until the original sending position is restored to service.

If the LF N combination at the end of the message fails to check, the sending position alarm signal is operated as described and the switch-to-supervisor signal is lighted at the receiving position. In this case the attendant at the receiving position sets the tape back to the start of the message and switches it manually to the supervisor, who resends it designated as a "suspected duplicate," as a complete or nearly complete message may have been transmitted before the trouble occurred. The attendant at the sending position operates a transmit-cancellation-notice switch to cancel any partial message that may have been recorded in the tape at that position.

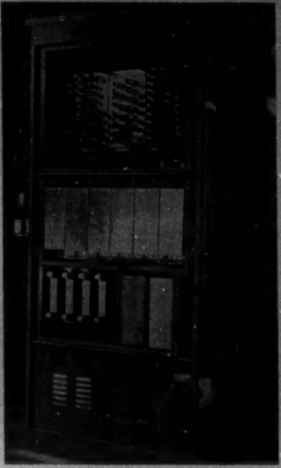


Fig. 9. Receiving cabinet, rear view, with electronic units at top

Multichannel Destinations

To handle heavy traffic loads to a particular destination, several outgoing line circuits and sending positions may be required. All sending positions in such a multichannel group represent the same destination. Any message for that destination may be switched to any sending position in the group that is available for use at the moment. The selection of a particular position is controlled by a load-distributor circuit. This circuit gives first choice to any sending position that has a tight tape condition, since such positions have no other messages on hand for outgoing transmission. During busy periods, when no position in the group has tight tape, each sending position in the group is selected in rotation as cross-office connections are established, so that the traffic is evenly divided among the channels.

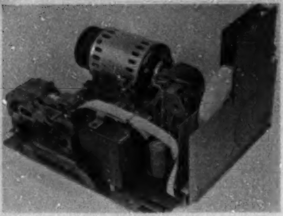


Fig. 10. Type 28 reperforator for receiving across office

Transmitter-Distributor

Each sending position is equipped with a Type 28 multicontact transmitter-distributor, shown in Fig. 11, for transmitting, to the outgoing line, the messages that the reperforator has recorded in perforated tape form. The transmitter-distributor operates at 60 or 100 WPM. It has two cam shafts; one, a transmitter shaft that drives mechanisms for feeding the tape and sensing the perforations, and the other a distributor shaft for sending character signals to the line in start-stop code. The shafts can operate together or separately, so that blank tape or other characters may be idled through the transmitter without sending to line and special character signals, controlled by a rotary switch, may be sent to line by the distributor contacts without stepping the tape through the transmitter.

When a message is received across office, slack tape allows the transmitter-distributor to start. At first, any blank tape is idled through the transmitter, with the distributor stopped, so that the blank characters are not sent to line. When the first character in the tape that is not a blank reaches the transmitter sensing pins, the transmitter shaft stops and the distributor seizes a connection to line; then both shafts operate to send the message to line. When the sensing contacts detect the EOM combination LF NNNN, both shafts stop and the line connection is released. These operations are repeated as each message is transmitted to the outgoing line.

Two-Position Outgoing Line Termination

Two sending positions may be arranged to send alternately to the same outgoing line circuit. Normally one position is used for high-precedence traffic and the other for low-precedence traffic. The transmitter-distributor of the low-precedence position releases the line at the end of each outgoing message to allow the high-precedence position to cut in, if it has a message to send. When this occurs, all high-precedence messages on hand are transmitted to line before the low-precedence position can seize the line again. This arrangement makes it possible to send a high-precedence message to line ahead of previously stored low-precedence messages. If the line is part of a multichannel group, the low-precedence position may be so arranged that it will accept and store not more than one message during periods when the high-precedence position is sending to the outgoing line. This is to avoid accumulating any considerable number of messages at the low-precedence position during periods when it cannot send to line.

Added EOM Characters

When an outgoing line is used as a trunk to certain other types of switching centers or to a multistation circuit (way wire), it may be necessary, for the proper operation of the other switching center or way station, to send certain added characters after the normal EOM. These added characters are controlled by setting a switch located inside the console. This switch has five positions as follows:

Position	Added EOM Characters	Purpose
1	none	sending to plan-55 centers or tributaries
2	5 blanks	sending to U.S. Army automatic switching centers
3	3 carriage returns 2 letter shifts	sending to Plan-51 centers
4	10 letter shifts	sending to torn-tape stations or centers
5	blanks, pauses	disconnecting a way station

Multistation Circuits: Way Wires

Multistation circuits extend from a switching center to two or three tributary stations. These circuits are operated duplex. Each station is equipped for selective control so that each message transmitted from the switching center is recorded only at the station to which it is directed. A selecting letter (X, Y or Z) is assigned to each station. This letter is transmitted by an automatic numbering machine immediately after the SOM characters ZCZC, and is recorded in the tape ahead of the message serial number. The upper sending positions are arranged to mount two or three numbering machines for this purpose, and to maintain an individual message number sequence with each station. As it passes through the transmitter-distributor, the selecting letter, in the tape, selects the proper calling signal so as to cut in the printer at the particular station that is to record the message. The printers at the other stations are locked out until the end of the message.

Fig. 11. Type 28 transmitter-distributor for sending to line

At each way station, messages are prepared in perforated tape form for transmission to the switching center. Only one station can send at a time. Each way station uses its own sequence numbers for transmitted messages. These numbers are checked automatically at the switching center by SNIs individual to each station. The SNIs are selected by selecting letters similar to those used for outgoing messages. The upper receiving positions in the incoming line cabinets are arranged to mount the additional SNIs. Depressing either a high-precedence or a low-precedence push button at a tributary station indicates, at the switching center, that a message is ready for transmission from a way station. The switching center equipment determines whether a high-precedence message is indicated and, if so, invites each station in turn with a high-precedence message to send, automatically accepting high-precedence messages as long as any are indicated. If there are none low-precedence messages are accepted in a similar manner. To give all stations an equal opportunity to send messages within the same precedence group, the invitation sequence always starts with the station following the one that has just previously transmitted a message.

Advantages of High-Speed Cross-Office Transmission

Messages are received over incoming lines at 60 or 100 WPM, are transmitted across office at 200 WPM and are then retransmitted to outgoing lines at 60 or 100 WPM. With the cross-office speed at least twice the speed of transmission to outgoing lines, enough messages may be sent across office to any sending position to keep the outgoing line busy, while the cross-office side of that position is idle at least half of the time. It is obvious therefore that the receiving positions will usually obtain an immediate cross-office connection and no appreciable amount of unsent tape should accumulate at these positions. If an occasional busy condition should cause an accumulation, however, the differential between the high cross-office speed and the lower incoming line speed would transmit messages away from the receiving positions at least twice as fast as they are being received, thus rapidly reducing the tape accumulation. The line circuits may thus be operated at a very high utilization level, without causing appreciable traffic delay. The transmitter-distributor at a line-sending position automatically functions to send continuously as long as there is any traffic at the sending position to send. Therefore the cross-office line need operate only intermittently to provide a full outgoing circuit load.

Switching to Busy Destinations

If a receiving position has a single-address message to be switched to an outgoing circuit that is busy, the message waits until a circuit to the desired destination becomes available. If the wait is unduly long, or if the message is a flash or emergency message, an alarm will operate to call an attendant to the position to take suitable action, such as interrupting a less important message or switching the message over an alternate route.

In the case of a multiple-address message, cross-office connections to the desired destinations are established one at a time. After a connection is established to the first available destination, any additional required connection that is delayed by a busy condition is diverted automatically to an intercept position after a timed interval of about 30 seconds. This is a standard sending position which has its outgoing line connected as an incoming line to an intercept receiving position, at which a copy of the message is stored and retransmitted to the required destination when a circuit becomes available.

When the required cross-office connections have been established, the numbering machine associated with each sending position selected transmits the next consecutive channel number for its circuit outlet. The SOM characters ZCZC are checked at each sending position and, if correct, the receiving position sends the incoming message number from the SNI and the precedence prosign characters from the tape to all connected sending positions. The routing indicators are then transmitted, sending each routing indicator to only one sending position, as required. The end-of-routing characters (CR CR LF D) are then broadcast to all connected sending positions, followed by the remainder of the message and the EOM characters LF NNNN. The cross-office circuits are then released.

If the ZC characters fail to check correctly at any of the connected sending positions, the message is not transmitted across office, but instead a number-cancellation notice is transmitted and the cross-office lines are released. As in the case of a single-address message, alarms are operated at the sending position in trouble and at the receiving position. An attendant called to the receiving position reswitches the message. If the EOM characters fail to check correctly, the message is resent as a suspected duplicate. Two copies of the message may be received at some destinations, but one will be designated "suspected duplicate."

Traffic Control Center

Although a switching center could be operated without a traffic control center, this equipment is normally provided, as it facilitates traffic handling and supervision. The traffic control center equipment includes a connection indicator board, a traffic routing board, a close-out indicator board, and receiving and sending printer sets.

Connection Indicator Board

The connection indicator board has a neon signal lamp and a push button for every receiving and sending position. It provides means for readily determining which sending and receiving positions are connected over cross-office lines at any time. Operating a push button on the board associated with any operating position lights the neon lamp for any operating position with which it is connected. This information is useful for maintenance purposes and also for certain operating conditions. For example, assume that a flash or emergency message has been received at a receiving position, and switched across office to some sending position. Such messages include bell signals which call an attendant to the receiving position. The attendant notifies a supervisor at the traffic control center that such a message is being transmitted from a particular receiving position. The supervisor depresses the push button corresponding to this position and ascertains the sending position to which it is connected. The supervisor then sends an attendant to that position to ensure that the message is not delayed by equipment or line trouble or by a large accumulation of message tape previously switched.

Traffic-Routing Board

The traffic-routing board is a jack board for making temporary changes in the routing of messages by the use of patching cords. Two pairs of jacks marked FROM and TO are provided for each destination to which this switching center is connected. One pair of jacks is for high-precedence traffic and the other for low-precedence traffic. Changes in routing are made by patching the FROM jack of one destination to the TO jack of another, thus diverting traffic from the first to the second destination. High- and low-precedence traffic may be independently routed. For example, if a trunk route between two switching centers becomes overloaded with traffic, or otherwise unusable, the high-precedence traffic or all traffic may be temporarily sent over an alternate route through a third switching center. By patching only the high-precedence jacks, high-precedence traffic for stations that are closed at night may be diverted to a separate position for commercial refile, while, at the sending position by simply stopping the transmitter, the low-precedence traffic is allowed to accumulate for later transmission when the station reopens.

Close-Out Indicator Board

The close-out indicator board contains groups of signal lamps to indicate the operating status of the sending line positions for the information of the supervisory personnel. These signals show which sending positions are closed out on the cross-office side or have their transmitters stopped for any reason.

Relay-Scanning-Design Technique Generates High Accuracy and Speed in Analog-to-Digital Transducer Measurements

ANDREW F. KAY
ASSOCIATE MEMBER AIEE

AN accelerated program of flight-testing an aircraft or its power unit it often instrumental in more speedy development of its final design for quantity production. The time rate at which flight-tests can be carried out is rarely dependent on a time-in-the-air factor, but instead hinges on the time required for calculating and computing the data resulting from individual tests. Tremendous amounts of data from many different sources must be reduced and correlated to permit equipment modifications before subsequent tests can take place. When data reduction of analog-type information to be used for computer calculations is performed by manual devices or human beings, the time between individual flight tests often spans a period of days, or even weeks. This does not take into account the increased chance for inaccuracies to occur and the extremely high cost of this type of data-reduction process.[1]

Simplified Data Reduction

Time and cost factors have stimulated a considerable amount of development in the area of analog-to-digital converters for data recording during the past few years. With these instruments, data can be taken directly from the source, converted to digital readout form, and recorded by various means such as punched or magnetic tape or telemeter directly to remotely located recorders. These digital recordings can be played back later in computer instruments without the need for further data reduction or conversion. By such means, final data computations are performed literally in a few hours so that subsequent flight tests can take place with a much shorter time interval. Thus a speedy day-after-day series of flight tests can greatly decrease aircraft or engine development time.

The problem of taking multiple transducer measurements becomes extremely difficult when the number of data-sensing elements increases. Great numbers of these sensing units require a data-recording system capable of handling a complete series of readings within a certain time cycle. In this case, data-recording-equipment size presents the greatest difficulty, particularly in aircraft applications. In addition to having data-reduction problems, analog recording equipment becomes extremely bulky when a large number of readings are necessary. Completely electronic digital recording devices, on the other hand, are extremely complex and large as well as being difficult to maintain because of adverse environmental conditions. With the factors of size, complexity, and ruggedness in mind, it was determined that a simplified relay scanning system of less than 1,000 cubic inches could be devised to function more effectively in high-accuracy digital reading of transducer measurements.

Discussion

RELAY SCANNING ANALOG-TO-DIGITAL CONVERTER

Some of the most recent techniques in analog-to-digital converter design include electromechanical high-speed scanning relays, chopper comparators, and elimination of preamplifiers. In operation, a system developed by Non-Linear Systems, Inc., which utilizes all of these techniques, Figs. 1 and 2, converts low-level d-c input signal voltages over a range of 20 microvolts to 20 millivolts into three decades of binary coded decimal pulses proportional to the input voltage. Each decimal unit is equivalent to 20 microvolts of the input voltage.

Basically, this analog-to-digital converter creates an internal voltage, termed the feedback voltage, which is varied until it equals the unknown input voltage from the transducer. As the instrument scans, the internal voltage becomes equal to this input voltage and as the system balances, the converter supplies an electric serial pulse train output to indicate the transducer voltage reading in digital form.

LOW-LEVEL AMPLIFIER VERSUS PREAMPLIFIER

Rather than using a preamplifier which raises the input voltage to a higher level before comparison, it was determined that more accurate and simplified comparisons could be reached by making comparisons at the actual level of the input signal. The error voltage is then amplified in the low-level amplifier which is within a digital servo loop. In this way not only is a separate preamplifier unnecessary, but also possible drift and distortion, often encountered in preamplifiers and their feedback networks, are eliminated.

Requirements for the low-level or error amplifier are much fewer than for a preamplifier. In this type of application, preamplifier gain must be held within 0.15% or better, and requirements for stability and phase response are even more critical. By comparison, a 20 or 30% variation in error amplifier gain will not affect the final reading appreciably. By using a chopper to sample alternately the reference and unknown voltages, certain advantages are obtained which make the low-voltage-level comparison possible. Choppers are available with noise levels held below 1/10 microvolt, which suits them particularly well for their application in this system. A further advantage of the chopper is evident in the fact that through metal-to-metal contacts there is no distortion such as might result if electronic switching with vacuum tubes or transistors were used.

AMPLIFIER

The d-c amplifier contains four high-gain amplifying stages as well as a restorer stage to maintain a constant reference level in the low-level amplifier. In addition, the amplifier circuitry includes two multivibrators; one actuates the restorer through a diode gate, while the other generates 400-cycle clock pulses. Receiving a square wave input from the low-level amplifier stages, the amplifier differentiates and clips this input to form

---

Paper 57-675, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 1, 1957; made available for printing September 19, 1957.

ANDREW F. KAY is with Non-Linear Systems, Inc., Del Mar, Calif.

Fig. 1. Exterior view of relay scanning analog-to-digital converter

a clipped sawtooth wave. This waveform is gated so that the clock pulse occurs coincidentally with the sawtooth trailing edge. Referred to as an anticlock pulse, the sawtooth cancels the effect which a lone clock pulse would have on a scanner unit. The scanner unit contains a bistable multivibrator which operates the high-speed scanning relay, which in turn connects and disconnects precision resistors in the reference voltage divider. Final output digits of the converter will be read in terms of the absence of pulses in the pulse train. Multivibrators in the amplifier circuit shape the clock pulses and pulse outputs for 1/800-second duration. An additional multivibrator provides reset pulses to reset all scanner units to starting position at the end of each digital readout period.

When the error voltage from the chopper switch is a positive square wave, indicating that the reference voltage is lower than the unknown voltage, the amplifier puts out an anticlock pulse. This anticlock pulse from the amplifier causes the digit being measured to readout. When the error is negative the amplifier has no output, and the clock pulse causes the scanner unit to have no readout.

CONVERTER INPUT

Since this analog-to-digital converter system utilizes a transformer input to the amplifier, it is particularly well suited to strain-gage and other transducer voltage measurements. Strain-gage transducers almost invariably require a measuring instrument which is completely isolated from ground. This factor is important in other transducers such as thermocouples which measure temperatures directly from metallic surfaces. In addition to erroneous readings, an unisolated system could create phase lags and require costly installation procedures to reduce detrimental effects. This system can be used on any measuring device which transforms any transient or quasi-static physical information into low-level direct voltages. However, the system could be easily modified to digitize high-voltage levels also. When properly programmed the input signal can be supplied by any combination of transducers as long as their outputs vary over the same voltage range. An example of this varied input might occur in an aircraft test application where, in addition to strain-gage and thermocouple readings, it is desirable to receive information from inputs of autopilot control, speed, and aircraft heading. When it is necessary to telemeter test readings to remote recorders, a serial digitized signal to the recorder is ideal since normal distortion due to communication equipment and conditions has no effect on the finally recorded reading. This is true since the digital output is in the form of discrete pulses as compared with varying amplitude signals where exact voltage levels determine the value of the reading.

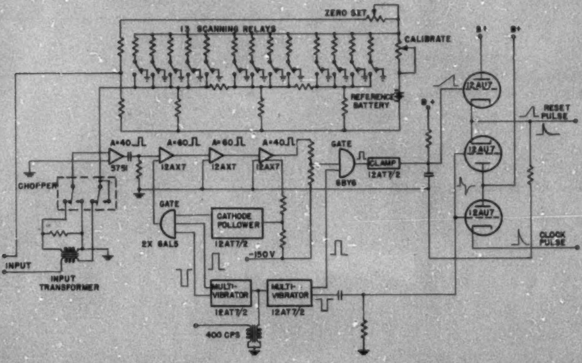

ZERO SET
13 SCANNING RELAYS
CALIBRATE
B+
B+
REFERENCE BATTERY
12AU7
12AU7
12AU7
RESET PULSE
CLOCK PULSE
A=40
A=60
A=60
A=40
5751
12AX7
12AX7
12AX7
GATE
6BY6
CLAMP
12AT7/2
CHOPPER
GATE
CATHODE FOLLOWER
12AT7/2
2X 6AL5
-150V
INPUT
INPUT TRANSFORMER
MULTI-VIBRATOR
12AT7/2
MULTI-VIBRATOR
12AT7/2
400 CPS

Fig. 2. Block diagram of converter design shown in Fig. 1

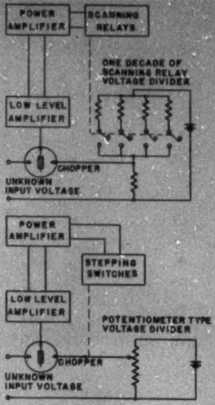

POWER AMPLIFIER
SCANNING RELAYS
ONE DECADE OF SCANNING RELAY VOLTAGE DIVIDER
LOW LEVEL AMPLIFIER
CHOPPER
UNKNOWN INPUT VOLTAGE
POWER AMPLIFIER
STEPPING SWITCHES
LOW LEVEL AMPLIFIER
POTENTIOMETER TYPE VOLTAGE DIVIDER
CHOPPER
UNKNOWN INPUT VOLTAGE

Fig. 3. Conventional voltage divider circuit

SCANNING

A chopper, vibrating at 400 cycles per second, alternately samples the unknown and the internal reference voltages. To the reference voltage a divider circuit of precision resistors is utilized. Connected to the known and calibrated internal voltage source, the divider circuit consists of groups of precision resistors which are alternately scanned. Rather than acting as a conventional voltage divider which samples a portion of a known voltage by means of a potentiometer, Fig. 3, this system combines various precision resistors in parallel which are in turn series-connected to a common resistance. The voltage generated in the common resistance is then the comparison reference voltage. As more resistors are connected to the common resistance, more current flows in the common resistance to raise the voltage to be compared directly with the unknown voltage.

High-speed hermetically sealed relays, which have a life expectancy of more than a billion cycles, sequentially scan the precision resistors by switching them in and out of a divider network. Scanning or sampling these resistors at a rate of 400 per second, the converter selects the combination of these resistors which equalizes the reference to the unknown input voltage. The position of the relays when the system is balanced determines the digital output reading. Identified by

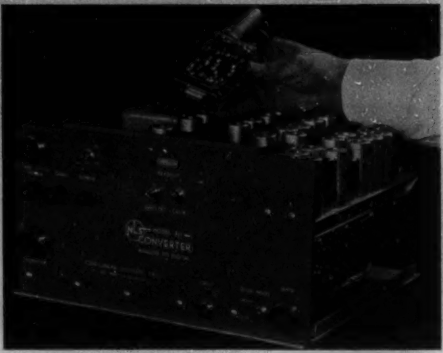


Fig. 4 (above). Scanner-unit modular construction

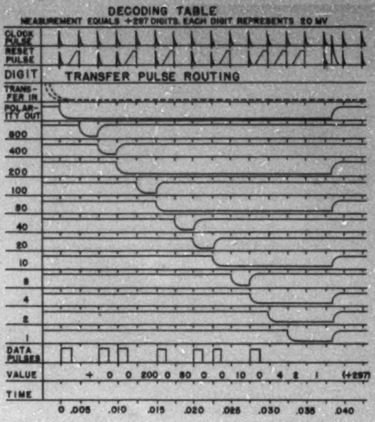


Fig. 5. Typical operational cycle of converter system

clock pulses provided by the amplifier circuit, the output is a serial train of binary coded data pulses.

The use of mechanical relays for switching the resistors during scanning is particularly advantageous when compared with electronic means such as vacuum-tube or diode devices. When the relay is open there is essentially infinite resistance between the contacts, and when it is closed there is essentially zero resistance. With electronic devices, in order to approach these optimum conditions, various complex compensating design techniques must be incorporated which add size and weight to the converter system.

Of the 13 scanner units, the first indicates polarity; the next four indicate hundreds of units; the following four, tens; and the last four are single units. Starting with the largest voltage digit, the scanner units increase the reference voltage until it equals the unknown voltage. Any time that the voltage digit or switched-in resistor makes the reference larger than the unknown, the voltage digit drops out. Each scanner unit is identical to the other 12, and is constructed as a completely interchangeable, modular, plug-in package; Fig. 4.

Operational Cycle

A typical operational cycle of this converter system is depicted in Fig. 5. Each time a clock pulse occurs, one resistor is sampled. If the increased voltage in the common resistance, which is caused by the addition of a sampled resistor, is greater than the unknown voltage, the amplifier signals the switching relay to reset the relay of this resistor. The next smallest resistor is then sampled in the same manner. Absence of a data pulse indicates that the digit is to be counted as a part of the total digital measurement. To obtain the voltage value of the digital measurement, the digits are added and their sum multiplied by 20 to obtain the reading in microvolts.

Coding

Generally, analog-to-digital converters read out in either 8-4-2-1 or 4-2-2-1 coding. Although this system can be adapted to either type of coding, the 4-2-2-1 readout was selected. With this coding, problems associated with redundancy are eliminated since the digits cannot add up to more than 9. This simplifies use of the taped output information in computers which generally incur difficulties if the decade sum is greater than 9. Decade totals with the 8-4-2-1 coding could, in some instances, be as high as 15.

Conclusions

The relay scanning analog-to-digital converter described provides one of the most favorable means for accurate and rapid recording of multiple voltage readings from low-voltage transducer sources. This system provides an optimum combination of currently available electronic and electromechanical components with respect to reliability and speed. Such innovations as mechanical divider circuit switching and elimination of preamplifier circuits have resulted in highly reliable, compact, and accurate systems for digitizing voltage measurements. The techniques described should pave the way for widespread future simplification of data reduction and computation for speedier product development in all fields of high-performance design.

Reference

1. Digital Computers Can Aid Utilities, F. J. Maginniss. *Electrical Engineering*, vol. 76, Feb. 1957, pp. 124–25.

Certain Applications of Matrices to Circuit Theory

LOUIS A. PIPES
MEMBER AIEE

Synopsis: This paper is intended to show the intimate relationships between the eigenvalues and eigenvectors of certain matrices and some well-known electrical parameters such as propagation constants, characteristic and iterative impedances in single-phase systems, and symmetrical components in polyphase systems whose impedance matrices have circular symmetry. The basic ideas are applied to show how Lucas functions, and the polynomials of Chebyshev and Gegenbauer enter naturally in the discussion of the propagation of potentials and currents along chains of identical and geometrically tapered quadripoles.

As IS NOW WELL-KNOWN, matrices provide a compact and flexible notation particularly useful in dealing with linear transformations, and presents an organized method for the solution of systems of linear differential equations.[1] The use of matrices in the treatment of problems in communication and electrical engineering has become increasingly popular. In recent years, it has been applied to the analysis and synthesis of 4-terminal networks[2] to the design of phase-shifting networks for oscillators,[3] to the theory of active quadripoles that include vacuum tubes as elements,[4] and to the analysis of polyphase power systems.[5]

The purpose of this paper is to emphasize that such well-known electrical parameters as propagation constants, characteristic and iterative impedances, and symmetrical components are intimately related to the eigenvalues and eigenvectors of certain square matrices.[6] The natural manner in which Lucas functions and the polynomials of Chebyshev and Gegenbauer arise in the analysis of chains of passive quadripoles is also demonstrated.

Transmission Matrix of Linear, Passive Quadripole

Consider the linear, passive quadripole or 4-terminal network of Fig. 1, which shows the general conventions of the directions usually adopted for the input and output potentials and currents. The 4-terminal network will be supposed to be an n-mesh network that has two accessible terminals as shown. The restrictions on the component elements of the internal structure of the network are that they be linear and constant. It is further assumed that the network is passive, i.e., the only sources of electromotive force are E_1 and E_2 produced by external agencies. As is well-known, the following linear relations exist between the complex potentials E_1, E_2, and the complex currents I_1, I_2 indicated in Fig. 1.[6]

$$E_1 = AE_2 + BI_2$$
$$I_1 = CE_2 + DI_2 \qquad (1)$$

The coefficients A, B, C, D are called the general 4-terminal network parameters in the electric circuit literature. It may be seen that, as a consequence of equation 1, A and D are dimensionless transfer ratios, the parameter B has the dimension of an impedance, and C the dimension of an admittance. If the elements of the internal structure of the quadripole are purely reactive, then the parameters A and D are real numbers and B and C are imaginary numbers. It may be shown that as a consequence of the reciprocity theorem,[6] the following relation exists between the four network parameters.

$$AD - BC = 1 \qquad (2)$$

To make the writing of subsequent equations more compact, it is convenient to introduce the column matrices or vectors x_1 and x_2 and the square matrix $[T]$ defined by the equations

$$\mathbf{x}_1 = \begin{bmatrix} E_1 \\ I_1 \end{bmatrix}, \; \mathbf{x}_2 = \begin{bmatrix} E_2 \\ I_2 \end{bmatrix}, \; [T] = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad (3)$$

With this notation, equation 1 and equation 2 may be written in the form

$$\mathbf{x}_1 = [T]\mathbf{x}_2 \qquad (4)$$

and

$$\det[T] = (AD - BC) = 1 \qquad (5)$$

where det = determinant.

The square matrix $[T]$ has been called the transmission matrix of the quadripole in the older literature of 4-terminal circuits. In certain recent papers, however, $[T]$ has been called the "chain matrix" of the quadripole.[7] Equation 4 expresses the fact that the two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are related to each other by a linear transformation. The relations between the four quantities E_1, I_1, E_2, I_2 given by equation 1 may be expressed in the following alternative ways:

$$\begin{bmatrix} E_2 \\ I_2 \end{bmatrix} = [T]^{-1} \begin{bmatrix} E_1 \\ I_1 \end{bmatrix}, \; [T]^{-1} = \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \qquad (6)$$

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = [Y] \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \; [Y] = \begin{bmatrix} D/B & -1/B \\ 1/B & -A/B \end{bmatrix} \qquad (7)$$

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = [Z] \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, \; [Z] = \begin{bmatrix} A/C & -1/C \\ 1/C & -D/C \end{bmatrix} \qquad (8)$$

$$\begin{bmatrix} I_1 \\ E_2 \end{bmatrix} = [g] \begin{bmatrix} E_1 \\ I_2 \end{bmatrix}, \; [g] = \begin{bmatrix} C/A & 1/A \\ 1/A & -B/A \end{bmatrix} \qquad (9)$$

$$\begin{bmatrix} E_1 \\ I_2 \end{bmatrix} = [h] \begin{bmatrix} I_1 \\ E_2 \end{bmatrix}, \; [h] = \begin{bmatrix} B/D & 1/D \\ 1/D & -C/D \end{bmatrix} \qquad (10)$$

In equation 7 and 8, the matrices $[Z]$ and $[Y]$ are defined to conform to the current and potential directions adopted in Fig. 1 and as a consequence of this, $[Z]$ and $[Y]$ are not symmetric matrices. The following relations between these square matrices are easily shown to exist.

$$[Z] = [Y]^{-1}, \quad \det[Z] = -B/C, \quad \det[Y] = -C/B \qquad (11)$$

$$[g] = [h]^{-1}, \quad \det[g] = -D/A, \quad \det[h] = -A/D \qquad (12)$$

The transmission or chain matrices of the more fundamental structures that occur in the theory of communication networks are given in Table I.

Eigenvalues and Eigenvectors of Transmission Matrix

Since the input and output quantities of the general quadripole of Fig. 1 are related to each other by the linear transformation equation 4, it is instructive to determine the eigenvalues and eigenvectors of the transmission or chain matrix $[T]$.[8] To determine these fundamental quantities, consider the linear transformation obtained by placing $\mathbf{x}_1 =$

Paper 58-52, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 21, 1957; made available for printing November 7, 1957.

Louis A. Pipes is with the University of California, Los Angeles, Calif., and consultant to the Hycor Division of the International Resistance Company, Los Angeles, Calif.

The author wishes to acknowledge his indebtedness to the Hycor Division of the International Resistance Company for sponsoring the research work that led to the writing of this paper, and to Ben Solow, Director of Research of the Hycor Division, for his encouragement and advice.

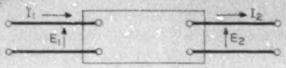


Fig. 1. Passive quadripole

x and $\mathbf{x}_2=\mu\mathbf{x}$ in equation 4. These substitutions give

$$[T]\mathbf{x}=\mu\mathbf{x} \qquad (13)$$

Let **x** have the elements x_1 and x_2, then equation 13 is equivalent to the two scalar equations

$$(A-\mu)x_1+Bx_2=0$$

$$Cx_1+(D-\mu)x_2=0 \qquad (14)$$

These two simultaneous equations give the following ratios:

$$x_1/x_2=B/(\mu-A)=(\mu-D)/C \qquad (15)$$

It may be seen that, as a consequence of equation 15, the parameter μ must satisfy the equation

$$\mu^2-\mu(A+D)+(AD-BC)=0 \qquad (16)$$

Since $\det[T]=(AD-BC)=1$, this equation reduces to the simpler form

$$\mu^2-\mu(A+D)+1=0 \qquad (17)$$

Equation 17 is the characteristic equation of the transmission matrix $[T]$.

Let $b=(A+D)/2=\text{trace}[T]/2$, where $\text{trace}[T]$ is defined as the sum of the principal diagonal elements of $[T]$. It is easy to show that the two roots of the characteristic equation 17 are given by

$$\mu_1=b+(b^2-1)^{1/2}$$

$$\mu_2=b-(b^2-1)^{1/2} \qquad (18)$$

The two roots μ_1 and μ_2 are the eigenvalues of the transmission matrix $[T]$. The eigenvalues have the property that

$$\mu_1\mu_2=1,\ (\mu_1+\mu_2)=2b=\text{trace}[T]=(A+D) \qquad (19)$$

Let

$$\mu_1=\epsilon^{a},\ \mu_2=\epsilon^{-a} \qquad (20)$$

then

$$(\mu_1+\mu_2)/2=(\epsilon^{a}+\epsilon^{-a})/2$$
$$=\cosh(a)=\text{trace}[T]/2=b \qquad (21)$$

The quantity

$$a=\cosh^{-1}(b)=1n(\mu_1) \qquad (22)$$

is recognized as the propagation constant of the quadripole.

It is noticed that if the eigenvalues μ_1 and μ_2 of the transmission matrix $[T]$ are distinct, and if $\mu=\mu_1$ and $\mu=\mu_2$ are substituted into equation 15, this determines two distinct ratios of x_1/x_2. These two ratios are the two eigenvectors of $[T]$. The two eigenvectors of $[T]$ are therefore

$$Z_1=(x_1/x_2)_{\mu=\mu_1}=(\mu_1-D)/C=\frac{(A-D)}{2C}+\frac{(b^2-1)^{1/2}}{C} \qquad (23)$$

$$=(A-D)/2C+\sinh(a)/C$$

$$Z_2=(x_1/x_2)_{\mu=\mu_2}=(\mu_2-D)/C=\frac{(A-D)}{2C}-\frac{(b^2-1)^{1/2}}{C} \qquad (24)$$

$$=(A-D)/2C-\sinh(a)/C$$

The quantities Z_1 and Z_2 are recognized to be the characteristic impedances of the transmission matrix $[T]$ are the characteristic impedances of the quadripole whose transmission or chain matrix is $[T]$. The iterative impedances of the quadripole are given by

$$Z_{1i}=\frac{\sinh(a)}{C}+\frac{(A-D)}{2C},\ Z_{2i}=\frac{\sinh(a)}{C}-\frac{(A-D)}{2C} \qquad (25)$$

From this it is seen that

$$Z_{1i}=Z_1 \text{ and } Z_{2i}=-Z_2 \qquad (26)$$

If the quadripole is symmetric, $A=D$ and relations 25 and 26 reduce to

$$Z_{1i}=Z_{2i}=Z_1=-Z_2=\frac{\sinh(a)}{C}$$
$$=(B/C)^{1/2}=Z_0 \qquad (27)$$

Powers of Transmission Matrix [T]

Since the over-all transmission matrix $[T]_n$ of a cascade of n identical quadripoles, each one of which has a transmission matrix $[T]$, is given[6] by

$$[T]_n=[T]^n \qquad (28)$$

it is convenient to develop a compact expression for integral powers of the transmission matrix $[T]$. The most fundamental manner of obtaining this desired expression is to make use of equation 13. If $\mathbf{x}_1$ and $\mathbf{x}_2$ be the eigenvectors that correspond to the eigenvalues μ_1 and μ_2, two equations are obtained of the form of equation 13:

$$[T]\mathbf{x}_1=\mu_1\mathbf{x}_1,\ [T]\mathbf{x}_2=\mu_2\mathbf{x}_2 \qquad (29)$$

If it is realized that only the ratios of the elements of the eigenvectors are specified by equation 15, the following may be written:

$$\mathbf{x}_1=\begin{bmatrix} Z_1 \\ 1 \end{bmatrix},\ \mathbf{x}_2=\begin{bmatrix} Z_2 \\ 1 \end{bmatrix} \qquad (30)$$

The square matrix

$$[M]=\begin{bmatrix} Z_1 & Z_2 \\ 1 & 1 \end{bmatrix} \qquad (31)$$

constructed by taking $\mathbf{x}_1$ and $\mathbf{x}_2$ as columns is the model matrix of $[T]$. In terms of the modal matrix, the two equations 29 may be combined into the single equation

$$[T][M]=[M]\begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}=[M][d] \qquad (32)$$

where $[d]$ is the diagonal matrix whose elements are the eigenvalues μ_1 and μ_2. This matrix has the following properties:

$$[d]^n=\begin{bmatrix} \mu_1^n & 0 \\ 0 & \mu_2^n \end{bmatrix},\ \text{trace}\,[d]=(\mu_1+\mu_2)$$
$$=2\cosh(a) \qquad (33)$$

If (4.5) is postmultiplied by $[M]^{-1}$, the result is

$$[T]=[M][d][M]^{-1} \qquad (34)$$

If equation 34 is multiplied by itself n times, the result is

$$[T]^n=[M][d]^n[M]^{-1} \qquad (35)$$

The inverse of the modal matrix $[M]$ is

$$[M]^{-1}=\frac{1}{(Z_1-Z_2)}\begin{bmatrix} 1 & -Z_2 \\ -1 & Z_1 \end{bmatrix} \qquad (36)$$

If the indicated matrix product of the right member of equation 35 is carried out, the result may be put into the following form:

$$[T]^n=\frac{1}{(Z_1-Z_2)}\times$$
$$\begin{bmatrix} (Z_1\mu_1^n-Z_2\mu_2^n) & -Z_1Z_2(\mu_1^n-\mu_2^n) \\ (\mu_1^n-\mu_2^n) & (-Z_2\mu_1^n+Z_1\mu_2^n) \end{bmatrix} \qquad (37)$$

If the values of the characteristic impedances given by equations 23 and 24 are substituted into equation 37, the result may be put into the convenient form:

$$[T]^n=(\mu_1^n-\mu_2^n)[T]/(\mu_1-\mu_2)-$$
$$(\mu_1^{n-1}-\mu_2^{n-1})[U]/(\mu_1-\mu_2) \qquad (38)$$

where $[U]$ is the unit matrix of the second order. This result expresses $[T]^n$ in terms of $[T]$ and its eigenvalues only. To write equation 38 in a more convenient form, let

$$\mu_1=\epsilon^{j\theta},\ \mu_2=\epsilon^{-j\theta} \qquad (39)$$

where $j=(-1)^{1/2}$.

Hence

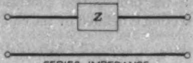
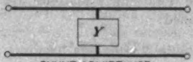
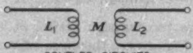
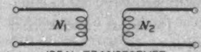
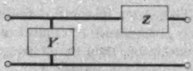
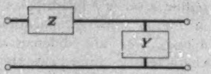
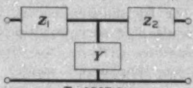
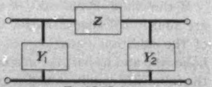
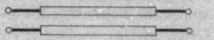
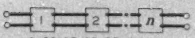
$$(\mu_1+\mu_2)=\epsilon^{j\theta}+\epsilon^{-j\theta}=2\cos(\theta)=\text{trace}\ [T] \qquad (40)$$

If these relations are substituted into equation 38 the result may be written as follows

$$[T]^n=\frac{\sin(n\theta)}{\sin(\theta)}[T]-\frac{\sin[(n-1)\theta]}{\sin(\theta)}[U] \qquad (41)$$

It is very convenient to express the relation 41 in terms of the Chebyshev

Table I. Transmission Matrices of Fundamental Quadripoles

No.	Network	Transmission Matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix} = [T]$
1	 SERIES IMPEDANCE	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$
2	 SHUNT ADMITTANCE	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
3	 COUPLED CIRCUITS	$\begin{bmatrix} \frac{L_1}{M} & \frac{jw}{M}(L_1L_2-M^2) \\ -j/wM & L_2/M \end{bmatrix}$
4	 IDEAL TRANSFORMER $a = N_2/N_1$	$\begin{bmatrix} 1/a & 0 \\ 0 & a \end{bmatrix}$
5		$\begin{bmatrix} 1 & Z \\ Y & (1+YZ) \end{bmatrix}$
6		$\begin{bmatrix} (1+ZY) & Z \\ Y & 1 \end{bmatrix}$
7	 T-SECTION	$\begin{bmatrix} (1+YZ_1) & (Z_1+Z_2+YZ_1Z_2) \\ Y & (1+YZ_2) \end{bmatrix}$
8	 π-SECTION	$\begin{bmatrix} (1+ZY_2) & Z \\ (Y_1+Y_2+ZY_1Y_2) & (1+ZY_1) \end{bmatrix}$
9	 TRANSMISSION LINE LENGTH = S	$\begin{bmatrix} \cosh(aS) & Z_0 \sinh(aS) \\ \sinh(aS)/Z_0 & \cosh(aS) \end{bmatrix}$ a = propagation constant Z_0 = characteristic impedance S = length of line
	 CASCADE OF IDENTICAL QUADRIPOLES $[T] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ = TRANSMISSION MATRIX OF EACH QUADRIPOLE OF THE CASCADE	Transmission matrix of entire cascade of n quadripoles $\begin{bmatrix} \cosh(an) & Z_0 \sinh(an) \\ \sinh(an)/Z_0 & \cosh(an) \end{bmatrix} = [T]_n$ Provided $A = D$ (symmetric networks) $a = \cosh^{-1} A$ = propagation constant $Z_0 = (B/C)^{1/2}$ = characteristic impedance $[T]_n = S_{n-1}(x)[T] - S_{n-2}(x)U,\ A \neq D$ $S_{n-1}(x)$ = Chebyshev polynomial $x = (A+D)$, U = second order unit matrix

polynomial $S_{n-1}(x)$.[9] This polynomial is defined by the expression

$$S_{n-1}(x) = \frac{\sin(n\theta)}{\sin(\theta)},\ x = 2\cos(\theta) = \text{trace } [T] \tag{42}$$

Thus it may be seen that equation 41 may be written as equation 43

$$[T]^n = S_{n-1}x[T] - S_{n-2}(x)[U] \tag{43}$$

where x = trace $[T]$.

This concise result is very useful and has been expressed differently by several writers. Parodi[10] expresses equation 41 in terms of the Lucas functions

$$U_n(x, 1) = (\mu_1^n - \mu_2^n)/(\mu_1 - \mu_2) \tag{44}$$

The result, equation 41, may also be expressed in terms of the Gegenbauer polynomials $C_{n-1}{}'(\cos\theta)$. These polynomials are defined by the equation[11]

$$C_{n-1}{}'(\cos\theta) = \frac{\sin(n\theta)}{\sin(\theta)} \tag{45}$$

In the critical case, when

$$b = \pm 1 = (A+D)/2 \tag{46}$$

the two eigenvalues μ_1 and μ_2 become equal to b. If $b=1$, $\theta=0$; while if $b=1$, $\theta=\pi$. Both of these values of θ lead to indeterminate forms in expression 41. If the indeterminate forms are evaluated, the following results are obtained:

$$\text{If } (A+D) = \ \ 2,\ [T]^n = n[T] - (n-1)U \tag{47}$$

$$\text{If } (A+D) = -2,\ [T]^n = n[T] + (n-1)U \tag{48}$$

The case of the symmetric quadripole arises frequently in practice. In this case $A = D$ and the propagation constant a and the characteristic impedances have the values

$$\begin{aligned} a = \cosh^{-1}(A),\ Z_1 = \frac{\sinh(a)}{C},\ Z_2 = -\frac{\sinh(a)}{C} \\ = Z_0,\ Z_2 = -Z_0 = -(B/C)^{1/2} \\ = (B/C)^{1/2} \end{aligned} \tag{49}$$

The eigenvalues of $[T]$ are given by equation 20. If these values for Z_1, Z_2, μ_1 and μ_2 are substituted into equation 37 the resulting expression reduces to

$$[T]^n = \begin{bmatrix} \cosh(an) & Z_0 \sinh(an) \\ \frac{\sinh(an)}{Z_0} & \cosh(an) \end{bmatrix} \tag{50}$$

This is the familar expression for powers of the transmission matrix of a symmetric quadripole.

Transmission Matrix of Geometrically Tapered Chain of Quadripoles

Most applications of equation 43 for the powers of a transmission matrix have been treated in the discussion of wave propagation along a uniform chain of quadripoles.[1,2,6] It will now be shown how the application of this expression may be extended to a cascade of geometrically tapered quadripoles. Consider the quadripole of Fig. 2. Let k be a positive real number. The transmission matrix of the quadripole of Fig. 2 is

$$\begin{aligned} [T]_n &= \begin{bmatrix} (1+YZ) & k^{2n}Z \\ k^{-2n}Y & 1 \end{bmatrix} \\ &= \begin{bmatrix} A & k^{2n}B \\ k^{-2n}C & D \end{bmatrix} \end{aligned} \tag{51}$$

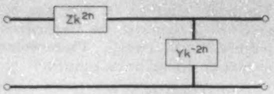


Fig. 2. Basic tapered quadripole

where, in this case, $A=(1+ZY)$, $B=Z$, $C=Y$, $D=1$.

If the index n is allowed to take on the values $n=0,1,2,3$, etc., a whole set of quadripoles of the type of Fig. 2 is obtained. Fig. 3 shows a chain of $(s+1)$ such quadripoles in which the index n has been allowed to take on the values, $n=0, 1,2,3,\ldots k$. The impedances and admittances of this chain of quadripoles are given by

$$Z_n=Zk^{2n},\quad Y_n=Yk^{-2n} \tag{52}$$

The transmission matrix $[T(s+1)]$ of the entire chain of $(s+1)$ quadripoles of the type of Fig. 2 is given by

$$[T(s+1)]=[T]_0[T]_1[T]_2[T]_3\ldots[T]_s \tag{53}$$

Now it is easy to see that the transmission matrix $[T]$ of the nth quadripole given by equation 51 may be factored in the form

$$[T]_n=\begin{bmatrix} k^n & 0 \\ 0 & k^{-n}\end{bmatrix}\begin{bmatrix} A & B \\ C & D\end{bmatrix}\begin{bmatrix} k^{-n} & 0 \\ 0 & k^n\end{bmatrix} \tag{54}$$

Let

$$[P]_n=\begin{bmatrix} k^n & 0 \\ 0 & k^{-n}\end{bmatrix},\quad [Q]_n=\begin{bmatrix} k^{-n} & 0 \\ 0 & k^n\end{bmatrix}=[P]_n^{-1} \tag{55}$$

Then, equation 54 may be written

$$[T]_n=[P]_n[T][Q]_n$$

where

$$[T]=\begin{bmatrix} A & B \\ C & D\end{bmatrix}=[T]_0 \tag{56}$$

The matrices $[P]_n$ and $[Q]_n$ have the following important property:

$$[P]_m[P]_n=[P]_{(m+n)},\quad [Q]_m[Q]_n=[Q]_{(m+n)} \tag{57}$$

Since $[P]_n=[Q]_n^{-1}$, it follows that

$$[P]_{n+1}=[P]_n[P]_1=[Q]_n^{-1}[P]_1 \tag{58}$$

Hence

$$[Q]_n[P]_{n+1}=[Q]_n[Q]_n^{-1}[P]_1=[P]_1 \tag{59}$$

The transmission matrix of the cascade of $(s+1)$ quadripoles may be obtained by substituting equation 56 into equation 53 and obtaining the following expression:

$$[T(s+1)]=([P]_0[T][Q]_0)([P]_1[T][Q]_1)\ldots([P]_s[T][Q]_s) \tag{60}$$

By the use of equation 59 this may be reduced to

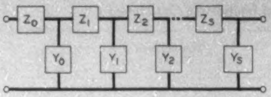


Fig. 3. Tapered chain of quadripoles

$$[T(s+1)]=([T][P]_1)^{s+1}[Q]_{s+1} \tag{61}$$

Now let

$$[T][P]_1=\begin{bmatrix} A & B \\ C & D\end{bmatrix}\begin{bmatrix} k & 0 \\ 0 & k^{-1}\end{bmatrix}=\begin{bmatrix} Ak & Bk^{-1} \\ Ck & Dk^{-1}\end{bmatrix}=[M] \tag{62}$$

It is then seen that the transmission matrix of the $(s+1)$ geometrically tapered quadripoles given by equation 60, may be written in the form

$$[T(s+1)]=[M]^{s+1}[Q]_{s+1} \tag{63}$$

Therefore, as a consequence of equation 43, the following may be written:

$$[T(s+1)]=S_s\mathbf{x}[M][Q]_{s+1}-S_{s-1}\mathbf{x}[Q]_{s+1} \tag{64}$$

where $S_s\mathbf{x}$ is the Chebyshev polynomial defined by equation 42, and

$$x=\text{trace }[M]=(Ak+Dk^{-1}) \tag{65}$$

This is a concise expression for the transmission matrix of the tapered chain of quadripoles of Fig. 3.

Chain of Symmetric Networks Having the Same Characteristic Impedance

As another example of a chain of nonuniform networks, consider the symmetric T network of Fig. 4. The transmission matrix of this network is

$$[T]_n=\begin{bmatrix} (1+Z_nY_n) & (2Z_n+Y_nZ_n^2) \\ Y_n & (1+Z_nY_n)\end{bmatrix}=\begin{bmatrix} A_n & B_n \\ C_n & A_n\end{bmatrix} \tag{66}$$

The characteristic impedance of this network Z_0 is given by

$$Z_0=(B_n/C_n)^{1/2}=(Z_n^2+2Z_n/Y_n)^{1/2} \tag{67}$$

Now, if Y_n is chosen, so that

$$Y_n=2Z_n/(Z_0^2-Z_n^2),\quad Z_0\neq Z_n \tag{68}$$

it is seen on substituting this expression into equation 67 that the characteristic impedance is Z_0 and is independent of the index n.

This choice of Y_n yields from equation 66, equation 69 following:

$$A_n=(1+Z_nY_n)=(Z_0^2+Z_n^2)/(Z_0^2-Z_n^2)$$
$$B_n=2Z_n+Y_nZ_n^2=2Z_nZ_0^2/(Z_0^2-Z_n^2)$$
$$C_n=Y_n=2Z_n/(Z_0^2-Z_n^2) \tag{69}$$

If the propagation constant a_n of this T section is defined by the equation

$$\cosh(a_n)=A_n=(Z_0^2+Z_n^2)/(Z_0^2-Z_n^2) \tag{70}$$

the following equations are easily obtained:

$$\sinh(a_n)=2Z_0Z_n/(Z_0^2-Z_n^2) \tag{71}$$

$$B_n=Z_0\sinh(a_n),\quad C_n=\sinh(a_n)/Z_0 \tag{72}$$

The transmission matrix of the network of Fig. 4 may therefore be written in the following form:

$$[T]_n=\begin{bmatrix} \cosh(a_n) & Z_0\sinh(a_n) \\ \dfrac{\sinh(a_n)}{Z_0} & \cosh(a_n)\end{bmatrix} \tag{73}$$

Therefore, if some definite value for Z_0 is chosen and Y_n is adjusted to satisfy equation 68, then by allowing Z_n to take the successive values $Z_1,Z_2,\ \ldots\ Z_r$, we obtain a set of r quadripoles of the type of Fig. 4 whose propagation constants $a_1,a_2,a_3,\ldots a_r$ are different but whose characteristic impedance is the same and is equal to Z_0.

The transmission matrix $[T(r)]$ of a cascade of r such quadripoles is given by

$$[T(r)]=[T]_1[T]_2[T]_3\ldots[T]_r \tag{74}$$

If the chain of matrices of equation 74 are multiplied together and use is made of the addition equations of the hyperbolic sine and hyperbolic cosine functions, the following concise result is obtained for transmission matrix of the cascade of r symmetric T networks of Fig. 4.

$$[T(r)]=\begin{bmatrix} \cosh(a_1+a_2+\ldots+a_r) & Z_0\sinh(a_1+a_2+\ldots+a_r) \\ \dfrac{\sinh}{Z_0}(a_1+a_2+\ldots+a_r) & \cosh(a_1+a_2+\ldots a_r)\end{bmatrix} \tag{75}$$

It is thus seen that the propagation constant $a(r)$ for the entire cascade is given by

$$a(r)=(a_1+a_2+\ldots+a_r) \tag{76}$$

where

$$a_n=\cosh^{-1}[(Z_0^2+Z_n^2)/(Z_0^2-Z_n^2)],\quad n=1,2,3,\ldots r \tag{77}$$

Hence equation 75 may be written concisely in the form

$$[T(r)]=\begin{bmatrix} \cosh a(r) & Z_0\sinh a(r) \\ \dfrac{\sinh a(r)}{Z_0} & \cosh a(r)\end{bmatrix} \tag{78}$$

Use of Matrices in Polyphase Systems

As is well known, matrix algebra is very useful in the analysis of certain classes of polyphase systems.[5] The intimate connection between certain transformations used by the power engineer and the eigenvalues and eigenvectors of the impedance matrices that arise in practical problems is not quite fully appreciated by many electrical engineers. In this section the basic ideas involved are presented as an extension of the analysis of the single-phase quadripole.

To fix the ideas in question, consider an apparatus with four terminals as shown in Fig. 5. Three of the terminals are to be connected to the lines of a 3-phase source and the fourth to the grounded neutral of the system. Let the complex line currents be I_1,I_2,I_3 and the complex line-to-ground voltages E_1,E_2,E_3 measured as shown in Fig. 5. By Kirchhoff's laws, the following equations may be written

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \tag{79}$$

or, more concisely

$$(E)=[Z](I) \tag{80}$$

In a very important class of electrical circuits which includes the equivalent circuits of 3-phase synchronous and induction machines with smooth air gaps,[12,13] the impedance matrix $[Z]$ of equation 80 has the general form

$$[Z]=\begin{bmatrix} Z_1 & Z_2 & Z_3 \\ Z_3 & Z_1 & Z_2 \\ Z_2 & Z_3 & Z_1 \end{bmatrix} \tag{81}$$

This matrix is symmetric about the second principal diagonal. A matrix that exhibits this type of symmetry is called a circulant matrix in the mathematical literature.[14] To determine the eigenvalues and eigenvectors of the matrix $[Z]$ it is necessary to follow the same procedure as that in the "Eigenvalues and Eigenvectors of Transmission Matrix" and to determine for what values of the parameter 9 the following relation is satisfied:

$$[Z]\mathbf{x}=\mu\mathbf{x} \tag{82}$$

where x is a column matrix or vector whose elements are x_1,x_2,x_3.

Equation 82 may be written in the form

$$([Z]-\mu[U])\mathbf{x}=(0) \tag{83}$$

where $[U]$ is the unit matrix of the third order and (0) is the third-order zero vector. Since equation 83 represents a homogeneous set of equations in the unknowns x_1,x_2,x_3, the following determinant must vanish:

$$\det([Z]-\mu[U]) = \begin{vmatrix} (Z_1-\mu) & Z_2 & Z_3 \\ Z_3 & (Z_1-\mu) & Z_2 \\ Z_2 & Z_3 & (Z_1-\mu) \end{vmatrix} = D = 0 \tag{84}$$

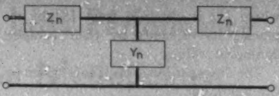


Fig. 4. Symmetric T-network

This is the characteristic equation of the impedance matrix $[Z]$; the possible values of μ satisfied by this equation are the eigenvalues of $[Z]$.

A determinant that exhibits the symmetry shown by D of equation 84 is called a circulant in the mathematical literature.[14] It is easy to show that the determinant D may be factored in the form

$$D=[(Z_1-\mu)+Z_2+Z_3][(Z_1-\mu)+a^2Z_2+aZ_3][(Z_1-\mu)+aZ_2+a^2Z_3] \tag{85}$$

where a = exponent $(2\pi j/3)$, the complex cube root of unity. The three eigenvalues of the circulant matrix $[Z]$ of equation 81 are therefore

$$\mu_1=(Z_1+Z_2+Z_3)$$
$$\mu_2=(Z_1+a^2Z_2+aZ_3),\quad a=\exp(2\pi j/3) \tag{86}$$
$$\mu_3=(Z_1+aZ_2+a^2Z_3)$$

It may be noted that the three eigenvalues have the dimensions of impedances.

To each eigenvalue μ_i there corresponds an eigenvector $\mathbf{x}_i$ which must satisfy an equation of the type of equation 82, or

$$[Z]\mathbf{x}_i=\mu_i\mathbf{x}_i,\quad i=1, 2, 3 \tag{87}$$

If the first element of each column $\mathbf{x}_i$ is chosen to be unity, then equation 87 fixes the other numbers of the columns and it may be seen that the three eigenvectors of $[Z]$ may be taken to be

$$\mathbf{x}_1=\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},\quad \mathbf{x}_2=\begin{bmatrix} 1 \\ a^2 \\ a \end{bmatrix},\quad \mathbf{x}_3=\begin{bmatrix} 1 \\ a \\ a^2 \end{bmatrix} \tag{88}$$

The square matrix $[S]$ formed from the three eigenvectors in the form

$$[S]=[\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3]=\begin{bmatrix} 1 & 1 & 1 \\ 1 & a^2 & a \\ 1 & a & a^2 \end{bmatrix} \tag{89}$$

is the modal matrix of the circulant impedance matrix $[Z]$. The three equations

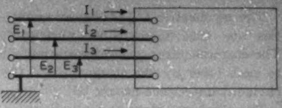


Fig. 5. Three-phase network

87 may now be written as the single matrix equation

$$[Z][S]=[S][d] \tag{90}$$

where $[d]$ is the diagonal matrix of the eigenvalues,

$$[d]=\begin{bmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{bmatrix} \tag{91}$$

It is thus seen that as a consequence of equation 90 the circulant matrix $[Z]$ may be written in the following diagonal form:

$$[S]^{-1}[Z][S]=[d] \tag{92}$$

The modal matrix $[S]$ is of great importance in the theory of symmetrical components[5]; it has the following important properties:

$$\det[S]=3(a-a^2)=3j\sqrt{3} \tag{93}$$

$$[S]^{-1}=\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix}=\frac{1}{3}[\bar{S}] \tag{94}$$

where $[\bar{S}]$ is obtained from $[S]$ by replacing each element by its complex conjugate.

In the theory of symmetrical components, the following linear transformations are introduced:

$$\mathbf{i}_s=[S]^{-1}(I)=\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix},\quad \mathbf{e}_s=[S]^{-1}(E)=\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{95}$$

The elements of the vectors $\mathbf{i}_s$ and $\mathbf{e}_s$ which are here denoted by small letters are the symmetrical components of the currents $\mathbf{I}$ and the potentials $\mathbf{E}$ respectively. As a consequence of the transformations, equation 95, the physical current and potential vectors are given by the inverse transformations in the form

$$\mathbf{I}=[S]\mathbf{i}_s,\quad \mathbf{E}=[S]\mathbf{e}_s \tag{96}$$

If these vectors are substituted into equation 80 the resulting equation is

$$[S]\mathbf{e}_s=[Z][S]\mathbf{i}_s \tag{97}$$

or

$$\mathbf{e}_s=([S]^{-1}[Z][S])\mathbf{i}_s=[d]\mathbf{i}_s \tag{98}$$

If this matrix equation is expanded, it may be seen that

$$e_1=\mu_1i_1,\quad e_2=\mu_2i_2,\quad e_3=\mu_3i_3 \tag{99}$$

The symmetrical component quantities are thus seen to be uncoupled. The corresponding quantities are related to each other by the eigenvalues of the circulant impedance matrix $[Z]$.

The 3-phase theory just discussed may be easily extended to n-phase systems whose impedance matrices exhibit circulant symmetry.[14]

Consider an n-phase network characterized by a circulant impedance matrix of the nth order of the form

$$[Z]_n=\begin{bmatrix} Z_1 & Z_2 & Z_3\ldots Z_n \\ Z_n & Z_1 & Z_2\ldots Z_{n-1} \\ \multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots\cdots} \\ Z_2 & Z_3 & Z_4\ldots Z_1 \end{bmatrix} \tag{100}$$

It may be shown that the eigenvalues of $[Z]_n$ are given by

$$\begin{aligned} \mu_1&=(Z_1+Z_2+Z_3+\ldots+Z_n) \\ \mu_2&=(Z_1+a^{n-1}Z_2+a^{n-2}Z_3+\ldots+aZ_n) \\ &\vdots \\ \mu_n&=(Z_1+aZ_2+a^2Z_3+\ldots+a^{n-1}Z_n) \end{aligned} \tag{101}$$

where in this case the number a is the complex nth root of unity given by

$$a=\exp(2\pi j/n) \tag{102}$$

where exp = exponent.

The modal matrix of $[Z]$ is given by the equation

$$[S]=\begin{bmatrix} 1 & 1 & 1\ldots\ldots1 \\ 1 & a^{n-1} & a^{n-2}\ldots a \\ \vdots & & \\ 1 & a & a^2\ldots\ldots a^{n-1} \end{bmatrix} \tag{103}$$

The typical element S_{pq} of $[S]$ may be written in the form

$$S_{pq}=a^{(1-p)(q-1)},\ a=\exp(2\pi j/n) \tag{104}$$

The inverse of $[S]$ is given by

$$[S]^{-1}=\frac{1}{n}[\bar{S}] \tag{105}$$

where $[\bar{S}]$ is obtained from $[S]$ by replacing its elements by their complex conjugate values.

The diagonal form of $[Z]_n$ is given by the following

$$[S]^{-1}[Z]_n[S]=[d_n] \tag{106}$$

where

$$[d_n]=\begin{bmatrix} \mu_1 & 0 & 0 & 0 & 0\ldots0 \\ 0 & \mu_2 & 0 & 0 & 0\ldots0 \\ \vdots & & & & \\ 0 & 0 & 0 & 0 & 0\ldots\mu_n \end{bmatrix} \tag{107}$$

If the current and potential vectors $\mathbf{I}$ and $\mathbf{E}$ of the nth phase polyphase system are related by the equation

$$\mathbf{E}=[Z](I) \tag{108}$$

Symmetrical component current and potential vectors are possible, as in equation 95 with the exception that $\mathbf{i}_s$ and $\mathbf{e}_s$ are now column matrices of n elements i_r and e_r, and the transformation matrix is now equation 105.

These linear transformations reduce equation 108 to the form

$$\mathbf{e}_s=([S]^{-1}[Z]_n[S])\mathbf{i}_s=[d_n](i)_s \tag{109}$$

and hence

$$e_r=\mu_r i_r,\ r=1, 2, 3, \ldots n \tag{110}$$

and the symmetrical component quantities are seen to be uncoupled.

The importance of transformations of the type in equation 95 in electrical engineering is due to the common occurrence of impedance matrices of the circulant type in practical power networks. Other types of linear transformations have been suggested and used to solve problems in which the impedance matrix has symmetry other than circulant symmetry.[15]

References

1. Modern Mathematics for the Engineer, **E. F. Bechenbach**, editor. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, chap. 13.

2. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-2, no. 2, June 1955. (This issue contains nine papers on the general subject of matrix methods of circuit analysis and synthesis and also an extensive bibliography.)

3. Design of RC Oscillator Phase Shifting Networks, **W. R. Hinton**. *Electronic Engineering*, London, England, vol. 22, no. 13, 1950, pp. 223–25.

4. Matrix Theory Applied to Thermionic Valve Circuits, **S. R. Deards**. *Ibid.*

5. Modern Electrical Engineering Mathematics (book), **S. A. Stigant**. Hutchinson's Scientific and Technical Publications, London, England, 1946. (Contains an excellent bibliography of the uses of matrices in power systems.)

6. The Matrix Theory of Four-Terminal Networks, **L. A. Pipes**. *Philosophical Magazine*, London, England, ser. 7, vol. 30, Nov. 1940, pp. 370–95.

7. On the Tabulation of Insertion Loss Low-Pass Chain Matrix Coefficients and Network Element Values, **S. D. Bedrosian, E. L. Luke, H. N. Putschi**. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XI, 1956, pp. 697–717.

8. The Mathematics of Physics and Chemistry (book), **H. Margenau, G. M. Murphy**. D. Van Nostrand Company, Inc., Princeton, N. J., second edition, 1956, chap. X.

9. Tables of Chebyshev Polynomials SN(x) and CN(x). National Bureau of Standards, Applied Mathematics Series 9, U. S. Government Printing Office, Washington, D. C., 1952.

10. Application des Polynomes Electrosphériques à l'Etude des Systèmes Oscillants a un grand Nombre de degrés de Liberté (book), **M. Parodi**. Gauthier-Villars, Paris, France, 1944.

11. Propagation des Ondes dans les Milieux Périodiques, **L. Brillouin, M. Parodi**. Masson et Cie., Paris, France, 1956.

12. Applications of the Method of Symmetrical Components (book), **W. V. Lyon**. McGraw-Hill Book Company, Inc., 1937.

13. Method of Symmetrical Co-ordinates Applied to the Solution of Polyphase Networks, **C. L. Fortescue**. *AIEE Transactions*, vol. 37, 1918, pp. 1027–1140.

14. Determinants and Matrices (book), **A. C. Aitken**. Oliver and Boyd, London, England, 1942, pp. 123–24.

15. Linear Transformations in Three-Phase Circuits, **Louis A. Pipes**. *AIEE Transactions*, vol. 60, June 1941, pp. 351–56.

Impedance and Return Loss Performance of Telephone Plant in Metropolitan Areas

L. B. BOGAN
MEMBER AIEE

THE FUNDAMENTAL DESIGN and successful operation of the nation-wide switching plan assumes a specified return loss performance of the metropolitan area plant. Metropolitan area plant includes the terminating trunks and subscriber loops which are involved at each end of an intercity telephone connection. The return losses not only must be adequate over the voice band as a whole in order to limit echo currents, but also must meet certain minimums at individual frequencies to avoid singing conditions.

The switching plan now in effect provides for more links and lower losses than the plan previously in effect. To operate at these lower losses with the same echo and singing margins, it is necessary to improve the terminal echo and singing return losses with respect to the conditions prevailing under the earlier plan.

This paper discusses the means of achieving the improved return losses, and reviews the limited test results which are now becoming available. First of all, however, it is thought that a better understanding of the situation will be realized if some attention is devoted to the nature of echo and singing and then to the peculiarities of the impedances of the metropolitan area plant as seen from the point of interconnection with the long intercity trunks.

Echo and Singing

It is never amiss to review the considerations controlling echo and singing. An understanding of the various features involved is fundamental for an appreciation of the techniques to be applied and the desired performance objectives.

Two factors contribute to the seriousness of an echo condition. One is magnitude of the returned, or echo, current and the other is its delay time. This paper will be concerned only with the factors contributing to the magnitude. The time factors of echo and the other criteria of intertoll trunk design are covered in other publications.[1,2]

Fig. 1, in its upper half, illustrates an over-all connection between two metropolitan areas separated by 1,000 miles. Two 500-mile carrier links, with a 4-wire switch at the intermediate office, are assumed to establish this connection. The objective loss design of this type of plant today is indicated by the db (decibel) quantities. In the lower half of Fig. 1 the significant features to observe are that echo is a single terminal phenomenon, and that singing requires the co-operation of both ends.

A low echo return loss at one terminal will promote echo to the far talker regardless of the magnitude of the echo return loss at the talker's own terminal. Also, it will be evident that as lower trunk losses become possible, the echo return losses at the terminals must be increased to maintain the same loss in the echo path.

Singing will occur when the return losses at the same frequency at both terminals approach zero and the net losses of the transmission medium, due to circuit variations, drift to gains. The significant factor here is that for a singing condition, the near-zero return loss at each terminal must be at the same frequency. It is possible to have a zero return loss at a low frequency at one terminal and a zero return loss at a somewhat higher frequency at the other terminal without singing jeopardy, even with the transmission path in a condition of gain.

Broadly speaking, return loss is a measure of the similarity or dissimilarity, depending on the point of view, of two impedances. The nearer the two impedances approach identity of magnitude and angle, the higher the return loss in decibels. Table I indicates the effects of magnitude and angular deviations individually on the return loss. The decibel values will provide a helpful background for the ensuing discussions.

For 2-wire switching systems the reference impedance condition for our terminal plant is the compromise network in the terminating sets of the intercity circuits. Generally, this is 900 ohms +2 mf (microfarads) for a crossbar tandem system, and 600 ohms +2 mf for the other types of 2-wire switching toll systems.

At 4-wire switching systems, where the hybrid coil to terminate the 4-wire intercity facility is permanently associated with the terminating trunk, compromise networks are not necessary. A precision network corresponding to the particular trunk facility may be used to obtain greatly improved return loss performance.

The problem in 2-wire switching is to make the heterogeneous array of impedances of the terminal plant somehow approach an impedance of 900 ohms+2 mf (or 600 ohms+2 mf) over the voice-frequency band.

As designed over the years, the return losses of the existing plant, against the compromise networks just referred to, appear to be a distribution having an average of 11 db with a sigma (standard deviation) of 3 db for the echo range. Singing points also appear to be a distribution, having an average of 6 and a sigma of 2 db. Tests made over the years have fairly well substantiated these figures. They have therefore been taken as the base values from which to work, if the singing and echo margins of the plant are to be maintained as the trunk losses are engineered for the proposed switching plan.

Within the past few years, one of the new techniques for the design of intercity trunks has been the omission of the 2-db pad loss in each terminating connection. The presence of this 2-db pad theoretically increases the average return loss of the terminating plant by 4 db (twice the loss of the pad). Consequently to satisfy the objectives with the new technique which omits this pad, or its equivalent loss, it is necessary to find a method of building an additional 4 db into the return loss of the terminating plant.

Another feature of the new approach in transmission design philosophy is the specification of performances in terms of distributions of various contributing factors rather than as a single "go, no-go" value. Methods of implementing this

---

Paper 58-4, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 7, 1957; made available for printing October 11, 1957.

L. B. BOGAN, deceased, was with the American Telephone and Telegraph Company, New York, N. Y.

A number of engineers in the operating telephone companies provided, and gave permission to use, test and survey data essential to the preparation of the paper. The author's associate, Miss Ruth L. Huxtable, contributed theoretical studies that were basic in the decision to adopt the improvement program. Important contributions were made by many other associates, particularly H. R. Huntley, O. F. Wallman, and H. H. Felder.

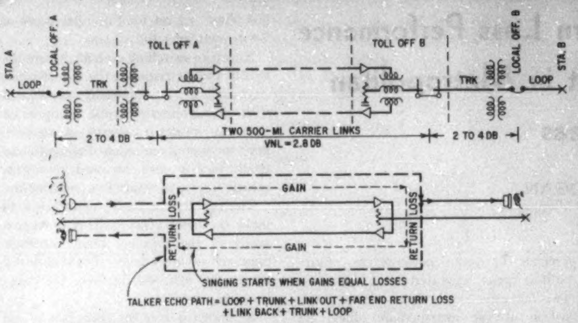

STA. A
LOCAL OFF. A
TOLL OFF A
TOLL OFF B
LOCAL OFF. B
STA. B
LOOP
TRK
TRK
LOOP
2 TO 4 DB
TWO 500-MI. CARRIER LINKS
VNL = 2.8 DB
2 TO 4 DB
GAIN
GAIN
RETURN LOSS
RETURN LOSS
SINGING STARTS WHEN GAINS EQUAL LOSSES
TALKER ECHO PATH = LOOP + TRUNK + LINK OUT + FAR END RETURN LOSS + LINK BACK + TRUNK + LOOP

Fig. 1. Over-all connection with example of losses, and development of talker echo and singing paths

concept in layout and maintenance are outstanding problems.

Loop Plant

The loop is identified as the channel extending from the serving central office to the subscriber's station, including the equipment permanently associated with the facilities at each end.

Loop facilities commonly encountered within a single central office area may be of 26, 24, and 22 gauge and, to a limited extent in older offices, 19 gauge. The lengths encountered obviously can start at practically zero for a station within the same building as the central office up to the maximum loop in the office. Loops up to 18 kf (kilofeet) are generally nonloaded, and those over 18 kf are loaded for the entire distance. Loaded loops present a somewhat different impedance condition from those to be presented in the following sections of this paper.

The variety of impedances in the loop plant, including the office battery supply circuit and the telephone set, are therefore due to the almost infinite variations resulting from the combination of lengths, gauges, and frequencies. Figs. 2 and 3 have been prepared to illustrate the variations within 24-gage nonloaded loops due to frequency and length.

Table I. Return Loss of Various Impedances Versus a Reference Impedance

Any Impedance in Terms of Reference Impedance	Return Loss In Db Versus Reference Impedance $Z\underline{/\theta}$	
$Z\underline{/\theta}$	∞	showing effect of variation in impedance magnitude only
$0.9Z\underline{/\theta}$ (or $1.11Z\underline{/\theta}$)	26	
$0.8Z\underline{/\theta}$ (or $1.25Z\underline{/\theta}$)	19	
$0.5Z\underline{/\theta}$ (or $2Z\underline{/\theta}$)	10	
$0.25Z\underline{/\theta}$ (or $4Z\underline{/\theta}$)	4	
$Z\underline{/\theta \pm 5°}$	27	showing effect of variation in impedance angle only
$Z\underline{/\theta \pm 10°}$	21	
$Z\underline{/\theta \pm 20°}$	15	
$Z\underline{/\theta \pm 45°}$	8	
$Z\underline{/\theta \pm 70°}$	3	

Fig. 2 shows four curves marked respectively, 0, 1, 10, and 20 kf. Each of these curves shows the effect on impedance as the frequency is changed from 300 to 3,000 cycles. The curve marked 0 kf is the variation over the frequency range of the input impedance of a battery supply repeating coil and a subscriber set with no intermediate line facility. In the other three curves these two terminations are separated by 1, 10, and 20 kf of 24 gage. Intermediate frequency points of 500, 1,000, and 2,000 cycles are indicated.

Fig. 3 shows the variation at three frequencies (300, 1,000, and 3,000 cycles) of impedances of 24-gage loops as they are extended from 0 to 20 kf. The impedance variations due to length and frequency of 24-gage alone blanket to a considerable extent the impedance field shown in the figures.

In Fig. 4, the 1,000-cycle impedance of 24-gage nonloaded facilities has been reproduced from Fig. 3 and, for comparison, curves for 19-, 22-, and 26-gage nonloaded loops from 0 to 20 kf are produced to suggest the range of the impedance. This is unduly simplified, since most of the loops have at least two gages and often more. In addition, in other than metropolitan areas, loop plant may include some bare copper or iron wire, involving other impedance contrasts. One problem currently being considered involves the poor impedance characteristics resulting from a combination of cable, open-wire, and magneto station sets.

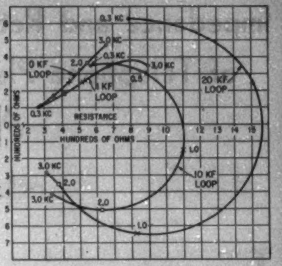


Fig. 2. Input impedance of 0-, 1-, 10-, and 20-kf 24-gage loops as the frequency is varied from 0.3 to 3.0 kc

Also to complicate the situation, the effect on impedance of substituting the 500-type set (the equalized set introduced as standard in the Bell System in the late 1940's,[3]) for the *F1A* set (the predecessor set which has been standard since about 1938[4]) on 24-gage is indicated. This shows rather radical impedance departure from the *F1A* set for short loop lengths. In future years, this change will be significant as the plant becomes saturated with 500-type sets. However, for the present, the problem is still mostly concerned with *F1A* sets since zoning specifies the use of 500-type sets on long loops where impedances fairly well conform to those shown with the *F1A* sets.

Trunk Plant

A terminal trunk is usually involved between an intercity circuit and the subscriber loop. This trunk is usually loaded and today *H88* loading (88-millihenry coils spaced at 6,000 feet) predominates. The gauge of the facility, however, may

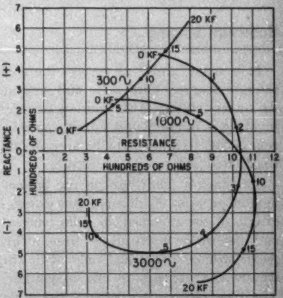


Fig. 3. Input impedance at 0.3, 1.0, and 3.0 kc as 24-gage loops are extended from 0 to 20 kf

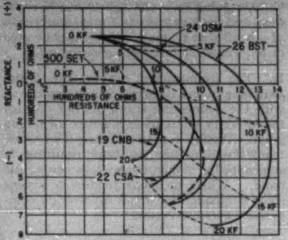


Fig. 4. Input impedance at 1.0 kc as 19-, 22-, 24-, and 26-gage loops are extended from 0 to 20 kf

be 19, 22, or 24, with possibility of 26 gage in some areas. Eventually a very large percentage of these trunks may be designed with repeaters of the negative impedance type.[5] The use of repeaters will add other complexities to terminal impedances. Also the use of carrier systems for these terminating trunks is becoming more common, but the present discussion will be concerned with only the voice-frequency facilities which predominate in the metropolitan plant.

The impedances of 19, 22, and 24 *H88*-loaded trunk facilities are not greatly dissimilar and the impedance of the 22 gage will serve also as a fair approximation of the other two.

Fig. 5 shows the mid-section impedance of a *22H88* pair and what happens to this input impedance when discontinuities or terminations are applied at a finite distance, corresponding to a practical trunk length.

Precision Network. This curve is the characteristic mid-section impedance of an infinite length of the facility. For present purposes this may be analyzed as a 24-kf trunk (four loading coils) terminated with the precision network for this type of facility properly adjusted for end section.

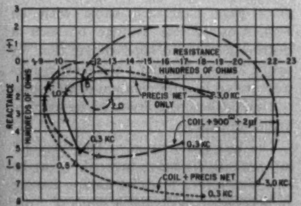


Fig. 5. Input impedance of 24-kf 22H88 trunk as frequency is varied from 0.3 to 3.0 kc, for three terminations other than loop

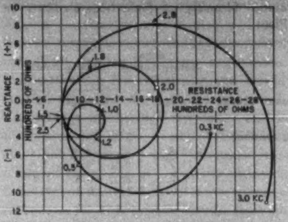


Fig. 6. Input impedance of 24-kf 22H88 trunk, terminated with a 20-kf loop, as frequency is varied from 0.3 to 3.0 kc

Coil Plus Precision Network. This curve shows the results of inserting a 1-to-1-ratio repeating coil between the 24-kf trunk and the precision network. It will be noted that a small loop is thrown into the characteristic at about 2 kc but that the largest effect on the impedance is at low frequencies, caused by the repeating coil. This curve also demonstrates that at the higher frequencies, the repeating coil is almost electrically transparent.

Coil Plus Compromise Network Termination. The precision network at the far end has been replaced with a compromise network of 900 ohms+2 mf, retaining the repeating coil. While there has been some additional low-frequency effect, and the loop has been enlarged, the major change in impedance has taken place at the frequencies above 2,000 cycles. At the higher frequencies the resistive component of the cable impedance is steadily rising whereas that of the termination remains unchanged. The junction irregularity increases with frequency.

Trunks Terminated in Loops

In Fig. 5, the sensitiveness of the input impedance of a trunk with any structural variation at a distant point is demonstrated. The irregularities assumed were comparatively nominal. Fig. 6 indicates the effect on the sending-end impedance of this trunk when the terminations beyond the repeating coil are replaced by a 20-kf loop of 24 gage. (At present, 18 kf is the maximum recommended nonloaded loop in the Bell System. The computations for 20 kf were available from an earlier study and are used in this instance as a fair approximation of a limiting nonloaded loop.) The relative scales of both resistance and reactance in this figure and subsequent figures are half those of Figs. 2 through 5.

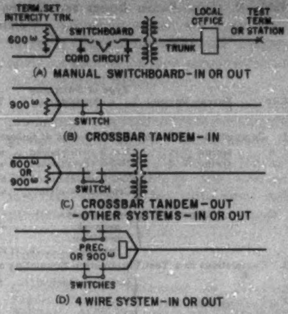


Fig. 8. Types of terminations at switching points

The gyrations in this curve are just short of fantastic. There appears to be a fair control of the impedance in the range of 500 to 2,500 cycles, but completely loose ends at 300 and 3,000 cycles. In this figure no repeating coil is involved at the input of the trunk. For plant conditions not requiring repeating coils this is substantially the best low-frequency impedance obtainable unless a low-frequency

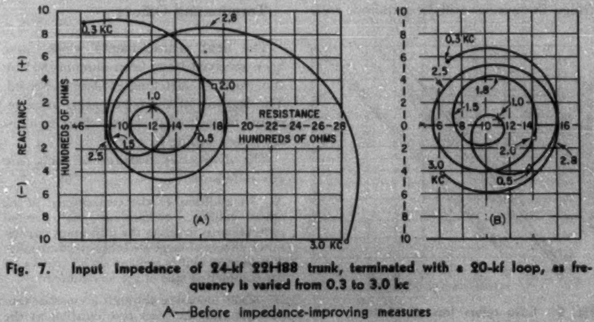


Fig. 7. Input impedance of 24-kf 22H88 trunk, terminated with a 20-kf loop, as frequency is varied from 0.3 to 3.0 kc

A—Before impedance-improving measures
B—After impedance-improving measures

Table II. Echo Return Losses and Singing Points: Summary of Field Measurements on Working Plant

City	Survey Period	Type of Switching System	Types of Trunks	Echo Return Losses in Db: Average	Echo Return Losses in Db: σ	Singing Points in Db: Average	Singing Points in Db: σ
1. Completions to Random Distributions of Subscriber Stations							
A	spring 1954	No. 3 switchboard	all	13.8	1.3	4.8	1.4
B	spring 1954	No. 3 switchboard	all	11.7	2.6	5.6	1.6
C	summer 1954	No. 3 switchboard	all	14.1	3.2	9.5	1.1
		Crossbar tandem	all	12.0	2.7	9.6	2.2
		4-wire compromise network	all	10.2	3.0	6.0	2.6
		precision network	all	22.7	2.9	17.0	2.9
2. Completions to a Test Termination Simulating an 8-Kf Loop of 24 Gage With F1A Set							
C	summer 1956	Crossbar tandem	*H88*				
			without IC	12.7	2.8	5.2	1.8
			with IC	16.5	3.1	6.8	1.7
			other than *H88*	11.3	1.8	3.3	1.5
		No. 3 switchboard	*H88*	11.1	2.2	3.5	1.0
3. Completions to a Test Termination of 900 Ohms + 2 Mf							
C	summer 1956	Crossbar tandem	*H88*				
			without IC	15.7	3.3	6.2	2.3
			with IC	19.1	3.8	8.0	2.2
			other than *H88*	11.9	2.0	4.2	2.1
		No. 3 switchboard	*H88*	12.4	2.4	4.3	1.0

corrector is used. This consists of a simple *RLC* series circuit tuned at about 250 cycles with a resistance of about 2,000 ohms shunted across the trunk input. With a low-frequency corrector the 300-cycle impedance would be reduced to the order of 1,000 ohms with a small angle.

All operating conditions, other than completing from crossbar tandems, involve a repeating coil at the input of the trunk. Traditionally, in the Bell System, the repeating coils have had mid-point capacitors of 4 mf in both windings.

In Fig. 7(A) an input repeating coil has been added to Fig. 6. Only the low-frequency impedances are affected, but it has been a major change. Fig. 7(A) is representative of the plant conditions producing the 11-db echo return loss and the 6-db singing point.

To improve the singing and echo performance of this type of impedance characteristic a radical operation on the fringe frequencies will be required, with perhaps only a nominal improvement over the body of the characteristic. The features found to produce these improvements were: 1. the substitution of a 1-mf capacitor for the 4-mf capacitor on the input side of the input repeating coil, and 2. the application of the impedance compensator (IC) technique to the loaded end section of the trunk.

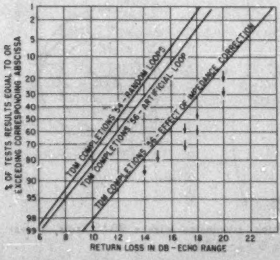


Fig. 9. Echo return loss distributions: incoming connections at crossbar tandem office

Fig. 7(B) shows the results of applying these two measures. Theoretically this appears to have achieved the desired improvements of echo and singing return losses by the wanted amounts.

Only actual measurements of modified plant layouts can determine to what extent the theoretical improvements developed in the foregoing have been realized. Fortunately, some field measurements are now becoming available; these are reviewed in the following section.

Switching Arrangements at Terminating Offices

The field measurements which have been received have been summarized by the class of switching system involved. The features of the terminating arrangements of each of these switching systems are shown in Fig. 8.

The distinctive features of each case are as follows:

A. In addition to the capacitance of the office cabling, completion via manual switchboards involves the important capacitance effects of large switchboard multiples. A repeating coil is present on all trunks on the completing side of the switchboard.

B. Calls incoming through a crossbar tandem provide the least complication at the terminating office.

C. Outgoing only through a crossbar tandem, and both IN and OUT connecting trunks at other 2-wire systems, such as step-by-step or no. 5 crossbar, require repeating coils.

D. The terminating set at the 4-wire switching point is permanently associated with the connecting trunk instead of with the intercity trunk as at a 2-wire switching point. Precision networks are used when trunk facilities are involved resulting in substantially higher return loss across the band. Where no trunk facility or a very short nonloaded facility is involved, a compromise network is used.

Summary of Field Measurements

Table II summarizes the results of echo return loss and singing point measurements made on working terminal plant.

Part 1 of this table shows, for the various types of switching systems, results of over-all measurements on random trunks and random subscriber stations in three different cities. The range of these results is comparatively narrow and conforms reasonably well to the anticipated distributions of echo and singing, with an average and sigma of the former of 11 and 3 db, and of the latter of 6 and 2 db respectively. Even the results for the 4-wire switching system with the precision network, if adjusted for the 10-db gain which may be applied, also fall in the middle of this range, becoming 12.7 db.

Parts 2 and 3 of Table II are measurements on random trunks, but only to two types of terminations, for the purpose of reviewing the over-all situation and setting up a background for specifying maintenance techniques. The problem is to find a termination available on a dial-up basis, which would provide the best reference, and the test objectives for this selected reference. Obviously, with a fixed termination the sigma should be less, though the average may be better or worse than the actual loop distributions, depending on the type of termination.

Part 2 tests were made to a termination consisting of an artificial loop of 1.5 miles of 24-gage nonloaded cable with an *F1A* subset. The average of such a distribution should approximate that for actual loops. In Part 3 the same trunks were tested to a termination of 900 ohms +2 mf. This was expected to average higher and did so.

The test termination of 900 ohms+2 mf currently is being favored, with the objective return loss distribution having an average of 18 db and a sigma of 2 db. However, this termination has some limitations unless all trunks appear at half-section impedance at the junction with the termination. With end sections of other than half-section impedance it will be necessary to work out different reference test values, which is undesirable.

Considerations are now being given to the relative costs and advantages of terminating all trunks at half-section. Increasingly wider use of the negative impedance repeaters which operate best with half-section cable impedance might make this a practical future technique.

The only comparative results currently available of the effect of using IC, one of the improvements already discussed, are shown under the crossbar tandem items in Parts 2 and 3 of Table II. Fig. 8(B) applies to these groups, because all were incoming through the tandem. The same trunks are not involved in comparative tests. Tests were made in the midst of a conversion program and different trunk groups are included in the classes of those with and without IC. Since only *H88* facilities are compared, there is no question that the approximately 4-db difference is a direct improvement resulting from the use of IC.

It will be noted that only a nominal singing improvement was obtained but this was expected since the IC alone does not affect the low-frequency impedance. Singing at the low-frequency end would control after the high end is improved. The change in the high-frequency end is illustrated by Figs. 6 and 7(B), while the low-frequency end remains unchanged as in Fig. 6. Further study is being given to the use of low-frequency correctors for this type of trunk, since no repeating coil is involved. Sufficient field measurements are not yet available to permit a report on the effectiveness of the improvements anticipated by the change in the capacitor in the repeating coil from 4 to 1 mf.

Fig. 9 shows the distributions of the echo return losses incoming at crossbar tandems at city C. Included for reference are the 1954 survey results as well as the 1956 comparative results.

The wide sigma for the 1956 results with the test loop termination is disturbing. Investigation shows it due mainly to a wide distribution of the averages of the groups, rather than to wide distributions within each group. To illustrate this, the arrow indicators under the right-hand curve of Fig. 9 represent the average point for each of the nine groups included in the over-all distribution.

Identification of the plant conditions producing the spread in results among the groups tested, and the improvement of the lower values to reduce the over-all spread, and incidentally the average, now lies ahead. This work may well produce the material for another paper.

References

1. NATION-WIDE SWITCHING OF INTERTOLL TRUNKS—TRANSMISSION CONSIDERATIONS, **O. F. Wallman.** *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 223–30.

2. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, **H. R. Huntley.** *Ibid.*, vol. 72, pt. I, Nov. 1953, pp. 670–76.

3. AN IMPROVED TELEPHONE SET, **A. H. Inglis, W. L. Tuffnell.** *Electrical Engineering*, vol. 70, Sept. 1951, pp. 770–75.

4. TRANSMISSION FEATURES OF THE NEW TELEPHONE SETS, **A. H. Inglis.** *AIEE Transactions*, vol. 57, Oct. 1938, pp. 606–12.

5. NEGATIVE-IMPEDANCE TELEPHONE REPEATERS—APPLICATION IN THE BELL SYSTEM, **Arthur F. Rose.** *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 430–35.

◆

Discussion

Arthur F. Rose (American Telephone and Telegraph Company, New York, N. Y.): Mr. Bogan's paper and the related papers[1,2] give an excellent picture of the problems involved in the establishment of nation-wide toll telephone service in the United States. The objective of adequate transmission between any two telephones in the country has been shown to be obtainable by proper design and good maintenance within the limitations imposed by economic considerations. I should like to emphasize some of the practical aspects involved in carrying out these objectives.

Mr. Bogan's paper deals with the terminal conditions and the steps required to meet the necessary requirements. It is important to point out that this is a co-operative matter, which means that the terminal office with poor impedance conditions does not suffer as much as the distant party whose echo is increased if the return loss objectives are not met. Smooth uniform terminal impedances are the result of careful checks and intensive activities to improve poor results, as may be seen by the spread between the cities shown in Fig. 10, a chart presented by Mr. Bogan at the time that his paper was read.

Engineering and design procedures which are under the control of the telephone company, such as attention to accurate load coil spacing, adjustment of compensators, and building out where necessary, are not enough. The telephone companies must deal with reliable manufacturers whose product is uniform and dependable. For example, there is little advantage in laying out the load coil sections to 2% accuracy if the inductance of the coils or if the capacity of the cable deviate from their nominal values by a much greater percentage than this.

Likewise, if telephone repeaters are used on the trunk circuits they should be of the type which does not introduce substantial irregularity. The advantages of the Western Electric Company's *E2-3* repeater in this respect was discussed by me in 1954.[3] The simple series-type of negative impedance repeaters are, of course, entirely unsatisfactory.

REFERENCES

1. See reference 1 of the paper.

2. STUDY OF RETURN LOSSES ON LOADED TRUNK CABLES AND METHODS TAKEN TO IMPROVE THEM, **G. H. Speake.** *AIEE Transactions*, vol. 77, pt. 1, May 1958, pp. 201–04.

3. See reference 5 of the paper.

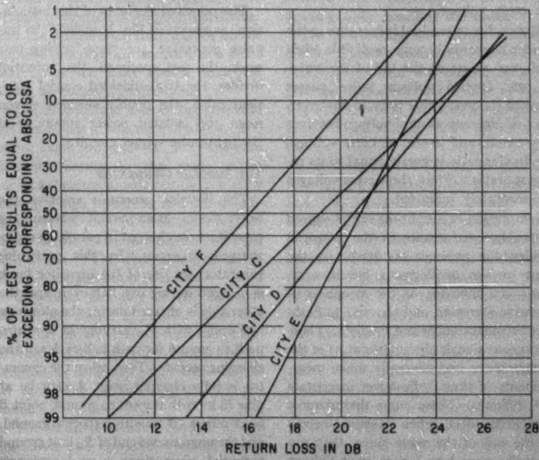


Fig. 10. Return losses for representative cities

Measurements of Steep-Front Impulse Waves with an Isolated Screen Room Installation

C. J. MILLER, JR.
FELLOW AIEE

J. F. WITTIBSCHLAGER
ASSOCIATE MEMBER AIEE

THE MEASUREMENT of steep-front impulse waves, when the time to flashover is less than 1 microsecond (μsec), presents problems seldom encountered in ordinary testing where flashover generally occurs on the tail of a conventional $1^1/_2\times$ 40 μsec wave.

In the fractional microsecond range, the voltage applied to the test specimen is generally increasing at a rapid rate at the instant flashover occurs. This means that the current through a resistor potential divider is likewise increasing at a corresponding rate, and hence the residual inductance in each element of this divider will generate an $L\,dI/dt$ voltage which is superimposed upon the IR voltage of the element. If the high and low sides of the resistor potential divider have different L/R ratios, the voltage division at the instant of flashover may be somewhat different than would be indicated by the resistance ratio alone. Thus, flashover voltages measured with such a divider may be somewhat in error, if only the resistance ratio is used in calculating the crest voltage.

Errors due to residual inductance in the divider elements become negligible when flashover occurs on the tail of the wave. As the applied voltage pulse passes through its crest value prior to flashover, the $L\,dI/dt$ component of voltage becomes very small, and hence, at crest voltage, the divider ratio is nearly equal to its resistance ratio. Thus, these crest voltages are accurately indicated.

In addition to amplitude errors caused by residual inductance in the potential divider, the presence of "hash" on the front of the oscillograms becomes increasingly annoying as the steepness of the wave increases, and the time to flashover becomes smaller and smaller. These extraneous signals obscure the start of the voltage wave and generally make measurements of time to flashover inaccurate and difficult. These same disturbances have little effect, when flashover occurs on the tail of the wave since they are generally gone long before crest voltage is reached.

The primary purpose of the doubly shielded isolated screen room which now houses the cathode-ray oscillograph (CRO) in the High Voltage Laboratory of The Ohio Brass Company at Barberton, Ohio, is to eliminate these extraneous signals.

The Isolated Screen Room Installation

The ideal place to locate the transient deflection plates of the CRO would be right across the low side of the potential divider with leads approaching zero in length. Since this location is generally quite impractical, the next best thing is to provide a circuit which will faithfully reproduce across the transient plates of the CRO the voltage which appears across the low side of the potential divider, without introducing any extraneous voltages from any other source. It is believed that the doubly shielded isolated screen room about to be described meets this requirement.

The principal elements of this installation, as shown in Fig. 1, include the impulse generator, the wave-forming network, the test specimen, the potential divider, the triple-shielded coaxial transient cable, the doubly shielded screen room, the isolated power supply, and the light-pulse trigger circuit.

The Impulse Generator

The impulse generator employs the conventional Marx circuit wherein the capacitors are charged in parallel and discharged in series. In this particular generator one side of the capacitor bank is charged during one half-cycle and the other side is charged during the alternate half-cycle, with R_{13} serving as a return path to ground for both half-cycles of the charging period. Thus, when the generator is fully charged point A may be at plus 75 kv with respect to ground, point B is at minus 75 kv with respect to ground, and the ungrounded end of R_{13} is at ground potential, since there is no longer any charging current flowing through it.

The impulse generator is fired by a discharge through the 3-ball gap, G_3 which is connected between point A and the common ground. Firing of G_3 causes point A to fall from plus 75 kv to ground potential. Because of the charge on the first two capacitors of the impulse generator, point B falls from minus 75 kv to minus 150 kv with respect to ground.

Since point D is momentarily held at its initial potential of plus 75 kv by the charge on the stabilizing capacitor C_6, the potential across the first series gap G_4 is suddenly increased from 150 kv to 225 kv. This causes gap G_4 to flash over, and thus connect point D to point B by a low impedance path.

The potential of point D then falls quickly from its initial value of plus 75 kv to minus 150 kv with respect to ground. This in turn places an excessive voltage across the second series gap, and the discharge process proceeds up the generator, eventually causing all series gaps to flash over and thus connect all capacitors of the generator in series.

Wave-Forming Network

The purpose of the wave-forming network is to control the shape of the wave applied to the test specimen. In the case about to be described, it was desired to apply a linearly rising wave to the test specimen, and to vary the rate of rise of this wave over a rather wide range of values. The connections and the range of magnitudes of controlling elements used in this particular investigation are shown in Fig. 1. The internal series resistance of the impulse generator is also considered to be a part of the wave-forming network, and was varied from 0 to 300 ohms during these tests. All resistances used in the wave-forming network are noninductive wire-wound units.

The Test Specimen

As shown in Fig. 1 the test specimen was a 12.5-cm (centimeter) sphere gap with a 6-cm spacing. This test specimen was chosen in order to demonstrate the effectiveness of the screen room and special potential divider by comparing test results with those obtained by Park and Cones[1] on a similar specimen using a similar potential divider at the National Bureau of Standards.

Paper 58-18, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 3-7, 1958. Manuscript submitted October 15, 1957; made available for printing October 28, 1957.

C. J. MILLER, JR., and J. F. WITTIBSCHLAGER are with The Ohio Brass Company, Barberton, Ohio.

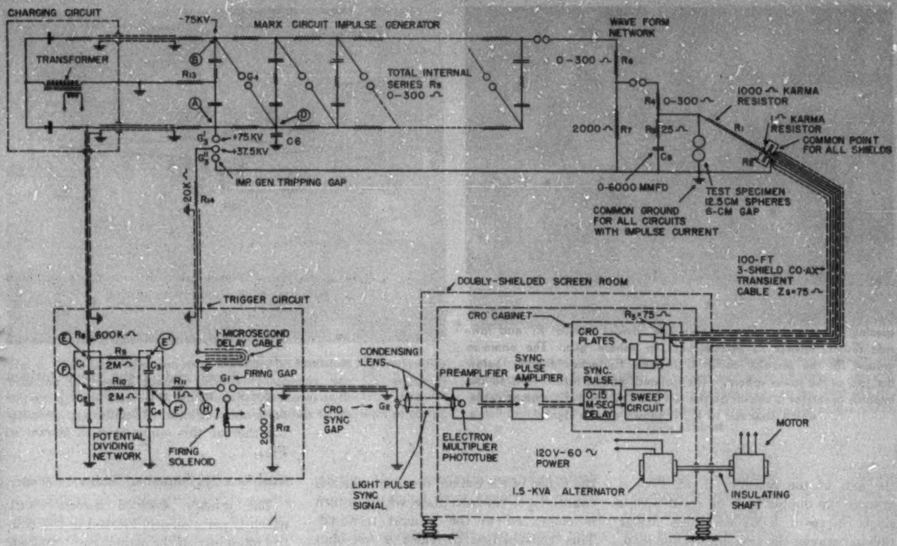

CHARGING CIRCUIT
TRANSFORMER
-75KV
MARX CIRCUIT IMPULSE GENERATOR
B
A
R13
G4
D
TOTAL INTERNAL SERIES Rs 0-300 Ω
G'3
+75 KV
+37.5KV
G''3
C6
IMP. GEN. TRIPPING GAP
20K Ω
R14
WAVE FORM NETWORK
0-300 Ω
R6
R4
0-300 Ω
1000 Ω KARMA RESISTOR
2000 Ω
R7
R5
25 Ω
R1
1 Ω KARMA RESISTOR
COMMON POINT FOR ALL SHIELDS
R2
C5
0-6000 MMFD
TEST SPECIMEN 12.5CM SPHERES 6-CM GAP
COMMON GROUND FOR ALL CIRCUITS WITH IMPULSE CURRENT
100-FT 3-SHIELD CO-AX TRANSIENT CABLE Zs=75 Ω
DOUBLY-SHIELDED SCREEN ROOM
TRIGGER CIRCUIT
R8
600K Ω
E
E'
1-MICROSECOND DELAY CABLE
F
C1
R9
2M Ω
C3
R10
2M Ω
C2
C4
R11
11 Ω
F'
H
G1
FIRING GAP
2000 Ω
R12
POTENTIAL DIVIDING NETWORK
FIRING SOLENOID
CONDENSING LENS
CRO SYNC GAP
G2
LIGHT PULSE SYNC SIGNAL
CRO CABINET
R3=75 Ω
CRO PLATES
PRE-AMPLIFIER
SYNC. PULSE AMPLIFIER
SYNC. PULSE
0-15 M-SEC DELAY
SWEEP CIRCUIT
ELECTRON MULTIPLIER PHOTOTUBE
120V-60 ~ POWER
1.5-KVA ALTERNATOR
MOTOR
INSULATING SHAFT

Fig. 1. The isolated screen room installation for the measurement of steep-front impulse waves

The common ground for the entire impulse circuit is at the base of the test specimen. The firing gap G_3 of the impulse generator is connected to this common ground. This method of ground keeps the main discharge currents out of the laboratory grounding grid and hence reduces differences in potential which might otherwise occur at different points in the grid if high level steep-front currents were intentionally introduced therein.

Even though a single common ground is used for the main impulse currents, there are charging currents which flow through the stray capacitances that exist between the laboratory ground grid and high-voltage portions of the impulse generator and its associated circuits. These charging currents flowing in the ground grid towards the single common ground point may produce sizeable potentials between different points in the grid, especially when they are produced by extreme steep-front voltage pulses.

The Potential Divider

The potential divider used in this investigation was patterned after the one described by Park and Cones.[1] Fig. 2 shows the divider connected across the 12.5-cm sphere gap with a 6-cm spacing. The top of the grounded sphere is approximately 13 inches above the rail which is intimately bonded to the laboratory ground grid. The point on the rail directly below the sphere is the common ground for all the impulse circuits. The ground connection of the potential divider and the ground lead from gap G_3 of the impulse generator are returned to a point just below the lower sphere, and just above the common ground point.

The high-voltage side of the potential divider R_1 has a resistance of about 1,000 ohms, and consists of ten 100-ohm Karma (Karma is the trade name for nickel-chromium-aluminum-iron alloy having a high resistivity and a low temperature coefficient of resistance) resistor units similar to those described by Park and Cones.[1] The over-all length of the divider is slightly over 50 inches.

Individual elements of the high-side divider are shown in Fig. 3. Each unit is made on a 1/2-inch-diameter polystyrene rod and is about 5 inches over-all length. A 3/8-inch-long brass ferrule at each end provides convenient terminals.

The resisitive element consists of two insulated and oppositely wound layers of Karma ribbon, 1/32 inch wide and 0.001 inch thick, separated by a double thickness of 0.001 inch thick by 0.50 inch wide Mylar (Mylar is the trade name for polyester insulating material) tape wound with half overlap. The second layer of Karma ribbon is also covered with Mylar tape to protect the winding while handling the units.

The low-voltage side of the potential divider R_2 has a resistance of about one ohm. This value was chosen so as to give the over-all divider a ratio of about 1,000 to 1, which provides a convenient magnitude of voltage on the plates of the CRO.

A close-up view of R_2 with the high side resistor R_1 connected to the line terminal and the triple shielded coaxial cable connected to the output terminal is shown in Fig. 4, with the shield removed. The purpose of the shield, which is shown in place in Fig. 2, is to prevent the possibility of any stray capacitance coupling between the junction of R_1 and R_2 and any external high-voltage circuit. With the shield in place the only current that can flow in R_2 is that which is conductively supplied to it through R_1.

Details of the construction of R_2 are shown in Fig. 5. The resistive element consists of four parallel loops of Karma ribbon, 1/16 inch wide by 0.001 inch thick, arranged 90 degrees apart about

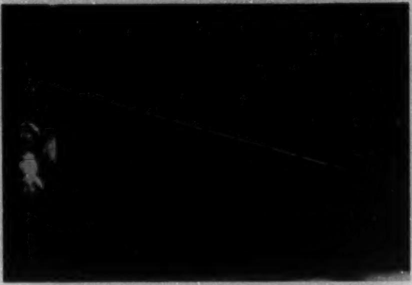


Fig. 2. Steep-front flashover of 12.5-cm sphere gap with 6-cm spacing showing location of high-side divider resistor R_1 and low-side divider resistor R_2 with respect to sphere gap. The common ground for the impulse circuit is located 13 inches directly below the top of the lower sphere. The isolated ground return lead to the impulse generator is shown on the left and the coaxial transient cable, which connects to R_2, is shown on the right

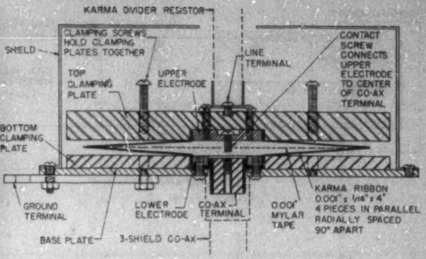

KARMA DIVIDER RESISTOR
SHIELD
CLAMPING SCREWS HOLD CLAMPING PLATES TOGETHER
TOP CLAMPING PLATE
UPPER ELECTRODE
LINE TERMINAL
CONTACT SCREW CONNECTS UPPER ELECTRODE TO CENTER OF CO-AX TERMINAL
BOTTOM CLAMPING PLATE
GROUND TERMINAL
LOWER ELECTRODE
CO-AX TERMINAL
0.001" MYLAR TAPE
KARMA RIBBON 0.001" x 1/16" x 4" 4 PIECES IN PARALLEL RADIALLY SPACED 90° APART
BASE PLATE
3-SHIELD CO-AX

Fig. 5. Details of the Karma 1-ohm resistor R_2 with clamping plates separated to show method of construction

the axis of the resistor. Each loop is arranged to double back on itself with a single layer of 0.001-inch-thick Mylar ribbon between the two sides of the loop. The ends of each loop are carefully soldered to circular electrodes. The four loops are firmly pressed between two clamping plates of insulating material. A standard coaxial terminal provides a convenient means for connecting to the high side of the 1-ohm resistor. Its center terminal is connected to the upper electrode by a contact screw. The upper electrode in turn is connected to the line terminal with four screws in parallel which provides a current path symmetrical about the axis of the unit. The lower electrode is intimately connected to the flange of the coaxial terminal which is connected to the base plate, which in turn is connected to the ground terminal. This construction provides a one-ohm resistor unit which has a very low inductance, and a convenient means for connecting to a coaxial cable.

THE TRIPLE-SHIELDED COAXIAL CABLE

The transient cable is basically an *RG-11/U* type with a surge impedance of 75 ohms. It is available commercially with a second outer shield insulated from the first shield, and also covered with an insulating jacket. A third shield was added by pulling a hollow braided sheath over the doubly shielded cable. A plastic coat was then pulled over the third shield, primarily for its mechanical protection.

The input end of the triple-shielded cable was fitted with a standard coaxial terminal. All three shields were bonded together at this connection as shown in Fig. 1.

THE DOUBLY SHIELDED SCREEN ROOM

The doubly shielded screen room, which is permanently located in the control room just off the main laboratory floor is shown in Fig. 6. While the control room is not completely electrostatically shielded from the field of the impulse generator, it does provide some degree of shielding for equipment and personnel.

The screen room is mounted on porcelain insulators so that it may be completely isolated from any ground connections in the control room.

As shown in Fig. 1, the outer screen is connected to the outer shield on the transient cable, the inner shield is connected to the middle shield on the transient cable while the CRO cabinet inside the screen room is connected to the inner shield of the transient cable. Thus, the CRO cabinet, the inner screen room, and the outer screen room form a series of three concentric Faraday cages within the innermost of which are located the transient plates of the CRO. The only common connection between the three concentric cages is located at the input end of the transient cable.

THE CRO

The CRO used in this investigation is a commercially available hot cathode type, which comes complete with sweep circuits, delay circuits, timing circuits, calibrating circuits, and all necessary power supplies for operation of all of these circuits. It is shown in Fig. 7 as it is normally used in the screen room.

For this investigation it was found that a very careful termination of the

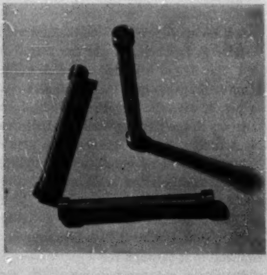


Fig. 3. A number of 100-ohm high-side divider resistor units showing the ferrule-type construction. Ten of these units are connected in series to make the 1,000-ohm high-side resistor, R_1

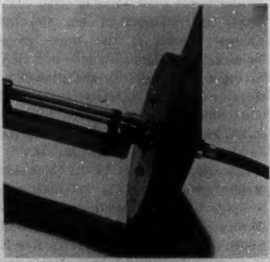


Fig. 4. Low side resistor R_2 with shield removed to show line terminal clip for connection to the high side resistor R_1. The four parallel resistance elements of R_2 are clamped between the two insulating plates which are shown to the left of the brass base plate

Fig. 6. The isolated screen room mounted on insulators as permanently installed in control room. CRO synchronizing gap G_2, which is mounted above motor with insulated shaft, is shown on the left. Impulse generator control cabinet and operator are shown to the right of screen room

Fig. 7. Interior of isolated screen room showing oscillograph and synchronizing pulse amplifier. Electron multiplier phototube and preamplifier cabinet with condensing lens is shown just to the right of CRO synchronizing gap G_2. The alternator is located directly below phototube cabinet

transient cable in its surge impedance is essential for successful operation with steep-front impulse waves. For this purpose a special low-inductance 75-ohm resistor R_3 was made by compressing in a zigzag fashion a suitable length of Karma ribbon with two layers of 0.001-inch Mylar tape between adjacent layers. In these tests the transient plates of the CRO were connected directly across the output end of the coaxial cable and its terminating resistor R_3.

SCREEN ROOM POWER SUPPLY

Since the interior of the screen room must be completely isolated from ground, it is essential that power for the CRO and other devices used therein be provided with a source which is completely isolated from ground. For this purpose a motor-generator set is used, with the motor located outside the screen room and the generator inside the inner screen. Complete isolation is provided by coupling the two units with an insulating shaft which extends through the walls of the screen room. The motor of this power supply is shown in Fig. 6 with its insulating shaft extending into the screen room.

SYNCHRONIZATION

In order to obtain suitable oscillograms of a transient phenomenon, it is necessary to provide a suitable sweep pulse on the horizontal plates of the CRO. Furthermore, it is essential that this sweep pulse be properly synchronized with the arrival of the transient pulse at its vertical plates. There are various methods for accomplishing this synchronization, all of which require a signal from somewhere in the system.

It is possible to bring an electric synchronizing pulse from a number of points in the impulse circuit, but all of these methods require the introduction of a second transient cable into the inner shielded cage of the CRO. This may lead to the introduction of spurious responses in the oscillograms.

LIGHT-PULSE SYNCHRONIZATION

In order to avoid the necessity of bringing a second transient cable into the screen room, a light-pulse synchronizing signal is used. This pulse of light, which travels readily through the walls of the screen room, cannot possibly introduce extraneous voltages inside the shielded cage.

The basic elements of the system are shown in Fig. 1. In the trigger circuit, equal capacitors C_1 and C_2 in series are charged to the same potential as the first capacitor in the impulse generator. Hence, point E in the trigger circuit has the same potential as point A during the charging period. The potential of point F is thus always at one-half the potential of point E with respect to ground.

Equal capacitors C_3 and C_4 are connected to C_1 and C_2 by two 2-megohm resistors. Hence, the potentials of points E' and F' are approximately at the potentials of points E and F respectively, since the time constants affecting the charging of C_3 and C_4 are short compared to the charging time of the impulse generator.

The middle sphere of the impulse generator tripping gap G_3 is connected to the point F' through a 20,000-ohm resistor R_{14}, a short connecting cable, a 1-μsec delay cable and an 11-ohm resistor R_{11}. Since practically no current flows through this path during the charging period, the middle sphere of gap G_3 is at the potential of point F' which in turn is approximately midway in potential between point E' and ground. Hence this middle sphere is approximately at one-half the potential of point A during the charging period and the gap G_3 does not flash over.

After the impulse generator is charged to the desired potential, the firing solenoid is energized and the midball of the firing gap G_1 is lifted mechanically. This causes gap G_1 to flash over, thereby permitting capacitor C_4 to discharge through R_{11}, gap G_1, and the 2,000-ohm resistor R_{12}. The rapid rise in potential across R_{12} causes the CRO synchronizing gap G_2 to flash over. Following this, capacitor C_4 discharges rapidly through R_{11}, and gaps G_1 and G_2.

After gap G_2 fires, point H falls rapidly to near zero potential, with R_{11} consuming most of the voltage still remaining on C_4. This drop in potential at point H is communicated after a one-μsec delay to the middle sphere of gap G_3. This causes G_3' to flash over and current to flow through R_{14}. Because of its high resistance, a small current through R_{14} causes the potential of the middle sphere to rise rapidly and thus impress the potential of point A across the lower portion of gap G_3. This causes G_3'' to flash over, and

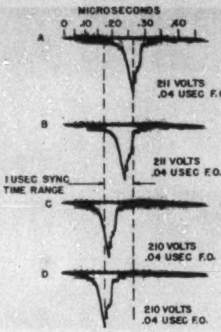


Fig. 8. Voltage oscillograms on a 12.5-cm sphere gap with 6-cm spacing with all connections as in Fig. 1, which includes the use of a light pulse for synchronization

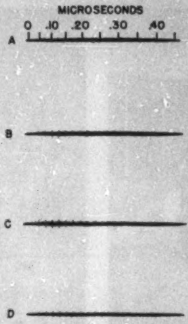


Fig. 9. Zero line oscillograms obtained with same conditions as those for Fig. 8, except lower end of R_1 is connected directly to ground terminal of R_2

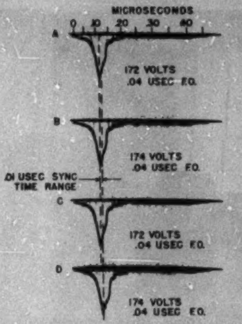


Fig. 11. Voltage oscillograms obtained with conditions same as those for Fig. 8, except a delay cable is used for synchronization

now since both G_3' and G_3'' are conducting, point A becomes connected to the common ground, and the impulse generator is discharged, as described in the preceding.

The flashover of gap G_2 produces a pulse of light which travels through the walls of the screen room to an electron multiplier phototube. A pair of condensing lenses focuses the arc in gap G_2 onto the phototube, thereby insuring a strong signal.

The output signal of the phototube passes through suitable amplifiers into the 0–15-μsec delay circuit which is provided in the CRO. The output of the delay circuit is introduced into the sweep circuit, where it triggers the sweep pulse.

In the circuit just described, the electric pulse created by the firing of G_2 initiates a train of events that ultimately culminates in the arrival of a transient signal at the vertical plates of the CRO. The light pulse, which is created simultaneously with the electric pulse, initiates a second train of events which ultimately culminates in the arrival of a sweep signal at the horizontal plates of the CRO. Suitable synchronization is effected when the sweep signal arrives at its plates just far enough ahead of the transient signal to establish a short zero line. Practically, synchronization is effected by suitable adjustment of the 0–15-μsec delay provided in the CRO.

Performance of the Installation

The performance of the over-all installation can be illustrated by comparing oscillograms taken under different conditions, with the same transient being applied, which in this case is the negative polarity 0.04-μsec flashover of a 12.5-cm sphere gap with a 6-cm spacing.

The four oscillograms shown in Fig. 8 are successive impulse shots applied to the test specimen. The freedom from "hash" on the front of the wave and the linear rise to flashover indicate the excellent results that may be obtained in the doubly shielded isolated screen room. This result is further emphasized by the lack of disturbances in the zero line oscillograms shown in Fig. 9 which were taken with the lower end of the high-voltage divider resistor R_1 connected to the ground terminal of R_2 rather than to its line terminal. The synchronization was the same for the oscillograms of both Figs. 8 and 9. If extraneous disturbances were present, this test would certainly reveal them.

Synchronization with the light pulse method appears to vary not more than 0.1 μsec for a series of repetitive shots as is shown by the four oscillograms in Fig. 8. This means that the two trains of events that are initiated by the firing of gap G_2 must be repetitive within this time interval.

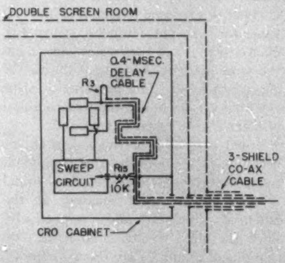


Fig. 10. Details of CRO synchronization using a delay cable

Delay Cable Synchronization

The use of a delay cable gives a method of synchronization which does not require a second incoming signal. As shown in Fig. 10, the incoming transient is used to trigger the sweep circuit, through a 10,000-ohm noninductive resistance tap at the end of the triple-shielded coaxial cable. The incoming signal must travel for an additional 0.4 μsec before it comes to the transient plates of the CRO. This delay provides sufficient time to get the sweep pulse started and thus provides synchronization.

Oscillograms obtained with delay cable synchronization are shown in Fig. 11. The results are equally as good as those obtained with light pulse synchronization shown in Fig. 8, but the amplitude is reduced from 210–211 volts to 172–174 volts, even though the magnitude at the source of these pulses is practically the same. The 18% loss in amplitude is due to attenuation in the 0.4 μsec delay cable. This means that if the delay cable were to be used for synchronization, a rather large correction factor would have to be used to compensate for cable attenuation, which could easily lead to inaccuracies. For this reason the delay cable method of synchronization does not appear as feasible as does the light pulse method.

Electric Pulse Synchronization

An external electric pulse may be brought into the screen room using

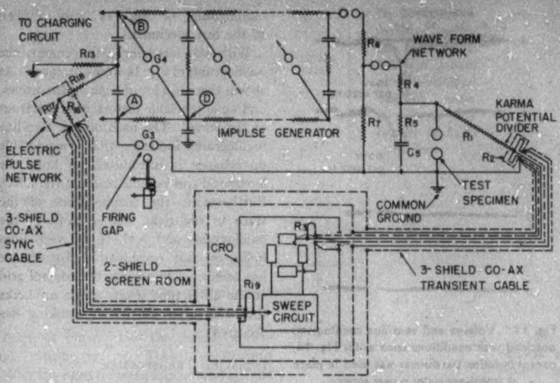


Fig. 12. Synchronization of CRO with electric pulse from impulse generator

a second triple-shielded cable. The diagram in Fig. 12 shows the details of one such method that was investigated.

The pulse is obtained from a potential divider network connected between the midpoint of the first capacitor bank of the impulse geneator and ground. While this point is practically at ground potential during the charging period, it suddenly jumps to minus 75 kv the instant gap G_3 is fired. A small fraction of this pulse, the magnitude of which is determined by the electric pulse network, is brought into the sweep circuit through a triple-shielded cable, which is terminated with a 75-ohm noninductive resistor R_{19}. The delay in breakdown of gap G_4 provides a suitable time lag between the starting of the sweep and the firing of the generator to accomplish synchronization.

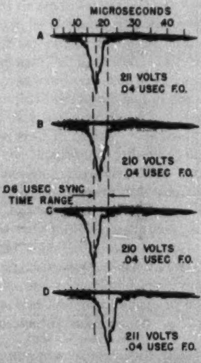


Fig. 13. Voltage oscillograms obtained with conditions same as those for Fig. 8, except an electric pulse is used for synchronization

In spite of the care used in shielding the cable and the input network, extraneous signals are introduced into the oscillograms as shown in Fig. 13. The source of these voltage oscillograms is the same as for the oscillograms of Fig. 8, and the amplitudes are about equal. However, there are superimposed oscillations on the front of the waves in Fig. 13 that are not present in Fig. 8. Apparently, these oscillations are caused by spurious signals, which are superimposed upon the sweep circuit rather than upon the transient plates. Here again the superior performance of the light pulse synchronization method is apparent.

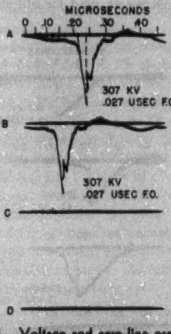


Fig. 14. Voltage and zero-line oscillograms on a 12.5-cm sphere gap with 6-cm spacing, with all connections as in Fig. 1 which includes screen room isolated from reference ground in control room. Zero-line oscillograms were obtained immediately following voltage oscillograms, but with lower end of R_1 connected directly to ground terminal of R_2. The same settings of the synchronization controls were used

Effect of Grounding Outer Screen Room in Control Room

When the impulse generator is fired, appreciable space charging currents may flow in the stray capacitances that inherently exist between the laboratory ground grid and the high voltage portions of the impulse generator and associated high-voltage circuits. Since all of the space currents that are intercepted by the ground grid must return through this grid to the common ground point at the base of the test specimen, it is possible that this common ground point may be raised in potential with respect to a reference ground which is outside the electrostatic field of the impulse generator.

When the screen room is connected only to the common ground at the base of the test specimen, as is shown in Fig. 1, it is possible for an appreciable difference in potential to exist between the screen room and grounded objects in the control room which are more or less shielded from the electrostatic field of the impulse generator. Oscillographic measurements show that peak voltages in the order of 35 kv exist between the screen room and the reference ground in the control room, when a very steep-front impulse voltage is applied to the 12.5-cm sphere gap with 6-cm spacing. While this pulse is of short duration, it does constitute a hazard to personnel in the control room, and it would be desirable to eliminate it, if it is possible to do so without affecting the resulting oscillograms.

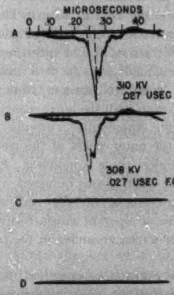


Fig. 15. Voltage and zero-line oscillograms obtained with conditions same as for Fig. 14, except outer shield is connected to reference ground in control room

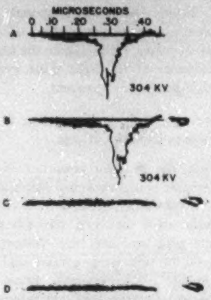


Fig. 16. Voltage and zero-line oscillograms obtained with conditions same as for Fig. 14, except oscillograph was placed outside of screen room and all three shields were connected to reference ground in control room

To investigate this possibility two sets of oscillograms were taken with identical pulses applied to the test specimen. The first set of oscillograms is shown in Fig. 14, which shows the results when the impulse generator, charged to 1,400 kv, is discharged into the test specimen. As shown, a crest potential of 307 kv is attained, and flashover occurred at 0.027 μsec. The front of the wave is practically linear and there is no hash at the start of the wave. The screen room is isolated, and synchronization is by light pulse. The two zero lines in Fig. 14 were made with R_1 connected to the ground terminal of R_2, thus bypassing the 1-ohm resistor R_2. There are no disturbances on the zero line, indicating complete freedom from stray pick-up.

The oscillograms shown in Fig. 15 were made under conditions identical to those of Fig. 14, except that the outer screen room was connected to the reference ground in the control room. The two flashover oscillograms are identical to those obtained in Fig. 14 with the screen room grounded, and the two zero lines indicate complete freedom from stray pick-up, even with the outer screen room grounded. Apparently, currents which flow in the outer shield of the transient cable under these conditions do not effect the oscillograms obtained. Thus, in the interests of safety to personnel, it is possible to operate satisfactorily with the outer screen grounded in the control room.

In the event it is suspected that the grounding of the outer screen room is influencing the oscillograms it is possible to take a set under both grounded and ungrounded conditions for checking this point. For this reason, it is believed desirable to maintain insulators under the screen room, so that this point may be checked occasionally.

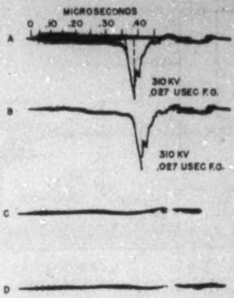


Fig. 17. Voltage and zero-line oscillograms obtained with conditions same as for Fig. 14, except isolation transformer was used in place of motor-generator power supply

CRO Removed from Screen Room

To emphasize the need for operating inside the screen room, the CRO was temporarily removed and placed on the floor just outside the screen room. The inner shield and central conductor of the triple-shielded cable were extended through the open door to the CRO and both this shield and the cabinet were grounded. The isolated power supply and light-pulse synchronization were used. The same steep-front impulse waves that were used for obtaining the oscillograms of Fig. 14 were applied to the test specimen.

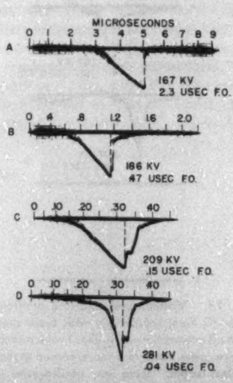


Fig. 18. Typical voltage oscillograms from steep-front tests on a 12.5-cm sphere gap with 6-cm spacing, using the installation described in Fig. 1. Volt-time data from these oscillograms are included in Fig. 21

With the foregoing arrangement, the oscillograms of Fig. 16 were obtained. As shown there is a lot of hash on the waves, and an appreciable distortion of the front of the wave. The hash on the zero-line oscillograms indicates that considerable disturbance is introduced on both the horizontal and vertical plates of the CRO. Furthermore, the reappearance of the trace to the right side of both of the oscillograms and both of the zero lines indicates that considerable disturbance has been introduced into the control grid of the CRO tube itself, which unblanks the beam after the initial blanking has occurred.

Isolating Transformer

To evaluate the advantage of the isolated power supply in the screen room, a test was made wherein all power to the screen room was supplied through a commercial isolating transformer instead of the motor-generator set. For this test, the CRO cabinet and both inner and outer screen rooms were connected to the reference ground in the control room. The test voltage applied to the specimen was the same as that used in obtaining the oscillograms of Fig. 14.

The results of this test are shown in Fig. 17. While the disturbances introduced by this isolating transformer are not as severe as those shown in Fig. 16, they do produce some disturbances which are particularly noticeable on the zero-line oscillograms. Unblanking of the

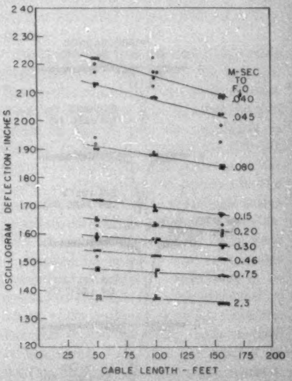


Fig. 19. Attenuation characteristics of triple-shielded transient cables of different lengths for voltages rising linearly to flashover at times indicated

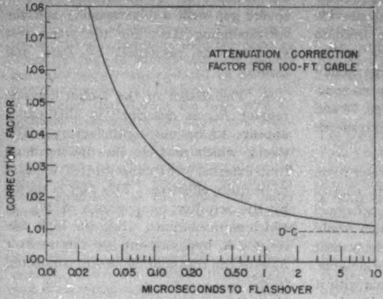


Fig. 20 (left). Attenuation correction factor for a 100-foot length of triple-shielded transient cable for voltages rising linearly to flashover

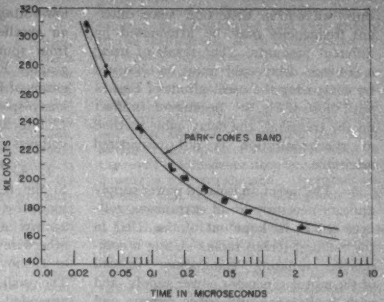


Fig. 21. The comparison of volt-time data for a 12.5-cm sphere gap with 6-cm spacing with negative polarity linearly rising voltage pulses applied to upper sphere, with a band obtained by Park and Cones on a similar test specimen. All voltages shown have been corrected to 25 degrees centigrade and 760 mm Hg

beam at the right of the oscillograms is also quite severe. These oscillograms show the need for complete isolation of the power supply as provided by the motor-generator set.

Correlation with National Bureau of Standards

To demonstrate the over-all performance of the complete installation as shown in Fig. 1, which includes the doubly shielded screen room, isolated power supply, light-pulse synchronization, triple-shielded transient cable, and Karma resistance divider, the volt-time curve of a 12.5-cm sphere gap with a 6-cm spacing was run with linearly rising voltages which produce flashovers in the time range of 0.027 to 2.3 μsec.

Typical oscillograms obtained during this series of tests are shown in Fig. 18 for times to flashover of 2.3, 0.47, 0.15, and 0.04 μsec. The corresponding oscillogram for a time to flashover of 0.027 μsec is shown in Fig. 14. As shown in all of these oscillograms, the fronts of these waves are practically linear clear up to the instant of flashover. The time to flashover was determined by extrapolating the slope of the upper half of the oscillogram back to the zero line.

The crest value was determined by scaling the maximum deflection on each oscillogram. Each deflection was then multiplied by the voltage sensitivity of the CRO tube, the attenuation correction factor of the triple-shielded 100-foot coaxial transient cable, the ratio of the Karma resistance divider, and finally the air density correction factor, to obtain the actual flashover voltage at standard atmospheric conditions.

ATTENUATION CORRECTION FACTOR OF TRANSIENT CABLE

Since the attenuation characteristics of the transient cable are undoubtedly different for waves of different steepness an investigation was made using three different lengths of cable for a number of different wave fronts. The results of this test, using cables of 50, 100, and 156 feet, are shown in Fig. 19. For each wave front, identical voltage waves were applied to the test specimen for each cable length. Hence, differences in oscillogram deflections are due only to the differences in attenuation in the different length cable. The repetitive nature of the pulses applied to the test specimen is indicated by the closeness of the plotted data points for a given cable length and steepness of wave.

As indicated in Fig. 19, straight lines with increasing slope at the shorter times to flashover fit the data quite well. From the slope of the line, the correction factor for a 100-foot cable may be determined. The curve shown in Fig. 20 gives this correction factor in terms of the time to flashover, and the slope of each line in Fig. 19 matches this curve for the appropriate time to flashover. Note that the curve of Fig. 20 becomes asymptotic to a horizontal line which corresponds to the d-c correction factor for the cable, based upon its resistance only.

CORRELATION WITH PARK AND CONES

Crest values of voltage, obtained by multiplying the oscillogram deflection by the sensitivity factor of the CRO tube, the cable attenuation correction factor from Fig. 20, the multiplying ratio of the Karma resistance divider, and the air density correction factor, are plotted in Fig. 21. Also plotted in this figure is the band given in Fig. 5 of the Park-Cones' paper from the National Bureau of Standards. As shown, there is an excellent correlation between results obtained using the equipment described in this paper and those reported by Park and Cones.[1]

Conclusions

1. The doubly shielded isolated screen room installation, including the triple-shielded transient cable, isolated power supply, and light-pulse synchronization, does an excellent job in preventing the occurrence of spurious response in the CRO.

2. One outstanding advantage of the light-pulse synchronization method is the complete freedom from extraneous voltages within the screen room. The effects of extraneous voltages that may be unintentionally introduced into the oscillograph, even with a carefully design circuit such as is shown in Fig. 12, are clearly indicated in the oscillograms of Fig. 13.

3. A second advantage of the light pulse synchronization method is the ease with which the transient may be positioned with respect to the sweep circuit, thus permitting it to be recorded in the most favorable position on the CRO screen.

4. While the use of a delay cable within the isolated screen room can provide synchronization without the introduction of a second signal, the 250-foot length required for satisfactory synchronization introduces considerable attenuation of the main crest and may also intro-

duce wave-form distortion since different frequencies may be attenuated by different amounts. The result of using a 0.4-μsec delay cable may be observed by comparing the oscillograms of Fig. 11 with those of Fig. 8. Because of its effect on the transient, the delay cable method of synchronization is not considered advisable.

5. The use of an isolated power supply appears necessary if all extraneous voltages are to be kept out of the CRO in the isolated screen room. Using a commercial isolating transformer instead of the motor-generator power supply, and with both inner and outer screens and CRO cabinet connected to the reference ground in the control room, the oscillograms of Fig. 17 were obtained. The presence of spurious responses, particularly of the zero-line oscillograms, indicates the unsatisfactory performance that may result if a completely isolated power supply is not used.

6. Zero-line oscillograms, which are taken when the lower end of the high-voltage divider resistor R_1 bypasses the low-voltage divider resistor R_2, provide an excellent indication of the freedom from spurious responses in the oscillograph. For example, the zero-line oscillograms of Fig. 9 show practically no spurious responses, while those in Figs. 16 and 17 indicate that appreciable extraneous voltages are present.

7. Connecting the outer screen room to the reference ground in the control room for the protection of personnel against high voltage pulses, which would otherwise exist momentarily at the instant of flashover, appears to be permissible. The oscillograms of Fig. 15, taken with it grounded, compare favorably with those of Fig. 14, taken with it isolated. In any event, as long as the insulators are kept under the screen room, it is possible to make such a check at intervals when other types of testing may be in progress.

8. The Karma resistance divider, which is described in this paper, appears to be quite satisfactory for steep-front impulse testing. This conclusion is based upon the excellent correlation obtained between the volt-time data on a 12.5-cm sphere gap with a 6-cm spacing and the corresponding data for the same test specimen, as presented by Park and Cones.[1]

9. The design of the 1-ohm low-side resistor R_2, as described in this paper appears to be quite satisfactory. The shield, which protects the line terminal from external stray capacitances, appears to be quite effective. This fact is shown by the zero-line oscillograms of Fig. 8, which were obtained when the high-side resistor R_1 bypassed the low-side resistor R_2 with the transient cable still connected to it.

10. The excellent correlation obtained with Park-Cones' data indicates the practicability of using the volt-time curve of a sphere gap as a reference standard for checking voltage dividers intended for steep-front measurements, as proposed by those authors.

Reference

1. Sphere Gap Volt-Time Curves—Reference Standards for Steep-Front Measurements, J. H. Park, H. N. Cones. *AIEE CP 57-215* (available on request).

Vol.
Pt. 1
1958

◆

Discussion

J. H. Park (National Bureau of Standards, Washington, D. C.): It is indeed gratifying when experimental results obtained independently in two laboratories agree almost perfectly as indicated by Fig. 21 in this paper. This is especially significant since agreement is obtained for the sparkover voltage of sphere gaps at times to sparkover down to 0.04 μsec. It definitely indicates the feasibility of using sphere-gap volt-time curves as reference standards for checking the accuracy of steep-front measurements.

However, before this curve can be accepted as part of a standard, one discrepancy must be resolved. As stated in the discusser's paper,[1] it would be expected that very slow rates of voltage rise would correspond to full wave sparkover so that for longer times to sparkover the volt-time curve should level off at the full wave sparkover value. At present this curve appears to be leveling off at about 150 kv which is 4.5% higher than the full wave value given in AIEE Standard No. 4 (143.5 kv). Data obtained last summer at the National Bureau of Standards (NBS) yielded a full wave sparkover value of 148 kv. Since this is very close to the minimum of the volt-time curve (as obtained both at NBS and Ohio Brass Company) it is suggested now that possibly the value given in AIEE Standard No. 4 is not correct for 12.5-cm spheres spaced 6 cm apart. As a check on full-wave measurements made at the NBS, data were also obtained on 25-cm-diameter spheres spaced 6 cm apart, giving a full wave 50% flashover voltage of 159 kv which agrees with the value obtained from Standard No. 4 to within 1 per cent.

The extreme shielding precautions found to be required at the Ohio Brass Laboratory indicate the importance of checking for stray pickup errors by the methods described in this paper. Of course it need not be concluded that such extreme shielding precautions must be taken at all laboratories. A continuously evacuated oscillograph whose deflecting system is inside a thick-wall iron cylinder, directly connected to the laboratory floor, would inherently have better shielding than a sealed glass envelope tube. An oscillograph with lower sensitivity would require a higher voltage from the divider low side and this tends to swamp out any pickup voltage. Also, the details of the exact method used to connect the center conductor and sheath of the measuring cable both at the divider low side and at the oscillograph are extremely important in eliminating pickup from ground currents. By making use of these factors, pickup troubles are practically eliminated at the NBS without resorting to double shielding of oscillograph and cable or isolation of power supply.

Reference

1. See reference 1 of the paper.

A. F. Rohlfs, J. Kresge, and **F. A. Fisher** (General Electric Company, Pittsfield, Mass.): This paper clearly illustrates the type of disturbances that may be encountered when attempting to measure short-duration, high-voltage impulse waves. It is interesting to note that the oscillograms show the signal to be measured with the correct amplitude even for those conditions where the wave shape is disturbed very badly. Evidently the disturbances were in the horizontal sweep circuits of the oscilloscope, rather than in the vertical deflection system. This is not too surprising since the horizontal sweep generator would have an output impedance much higher than the 75-ohm impedance level of the vertical deflection system.

We have encountered the same sort of problems and have observed sweep distortion phenomena very similar to those shown in the paper. Our approach toward eliminating such disturbances has been along two distinct lines. One has been to improve the grounding circuitry; the other has been to improve the oscilloscope circuitry so that it is less susceptible to external disturbances. This has been done by lowering the impedance of the horizontal sweep generator, isolating the vertical and horizontal deflection systems as much as possible and balancing the residual capacitances between the horizontal and vertical deflection systems. So far these approaches have been adequate.

There can be no question about the effectiveness of using the doubly shielded room, light-pulse synchronization, and isolated power supply. However, the question of personnel safety must be faced and in our judgment it requires that the oscilloscope chassis be grounded directly to the shielded room in which it is placed. If the oscilloscope were grounded in this manner, we wonder how much improvement could have been obtained if only the horizontal sweep circuit had been worked on?

There appears to be one apparent discrepancy between the caption for Fig. 17 and the body of the paper. The caption states that all conditions were the same as those for Fig. 14 except that the motor-generator supply was replaced by an isolation transformer. In the body of the paper it is stated that during the tests with an isolation transformer the inner and outer screened rooms were connected to the control room ground. If this is true then can any rigorous conclusion be drawn about the effectiveness of an isolation transformer when grounding of the inner and outer screened rooms may have had a detrimental effect in itself? Have the authors made tests using an isolation transformer but with the inner and outer screened rooms isolated from the control room ground and if so would they comment on the results?

C. J. Miller, Jr. and J. F. Wittibschlager: Mr. Park suggests the possibility that the negative flashover voltage of 143.5 kv for a 6-cm gap between 12.5-cm spheres, as now given in AIEE Standard No. 4, may be somewhat in error, since the volt-time curve which he obtained with impulse waves having a linearly-rising front levels off at about 150 kv at longer times to flashover. This possibility is further increased by observing a negative critical flashover voltage of 148 kv for this same gap, using the conventional $1^1/_2 \times 40$-μsec wave.

In order to check this value, a critical flashover test was run in Barberton, using a conventional $1^1/_2 \times 40$-microsecond wave on the 6-cm gap spacing and arrangement shown in Fig. 2. The results of this test gave a negative critical flashover value of 148.5 kv, which agrees very well with Mr. Park's test value of 148 kv. Thus, it would appear that the value of 143.5 kv given in AIEE Standard No. 4 is somewhat low. As a check on the accuracy of voltage measurement, a positive critical flashover value of 152 kv was obtained at the same time, and this value agrees quite well with the 153.8-kv value given in AIEE Standard No. 4. It would thus appear that the discrepancy indicated by Mr. Park may now be confirmed. However, confirmation from a third laboratory is highly desirable.

Messrs. Rohlfs, Kresge, and Fisher indicate that they, too, have encountered extraneous voltages in the horizontal sweep circuit which disturb the wave shape in a manner similar to that shown in the paper, and that they have been able to eliminate such disturbances by improving the grounding circuitry and the oscillograph circuitry. Such an approach appears to be impractical when using a commercially available hot-cathode type of oscillograph because of the intricate electronic circuits involved. It became a much simpler problem just to prevent the disturbances from ever getting to the oscillograph, as is accomplished by the doubly-shielded screen room.

As to personnel safety, the oscillograph chassis and the inner screen room are bonded together at the input end of the transient cable. The only potential which can exist between the chassis and inner screen at the operator's position would be the result of impedance drop due to currents flowing in the two sheaths. Operating experience has shown that this potential is negligibly small.

In running the test with an isolation transformer, the results of which are shown by the oscillograms in Fig. 17, it was necessary to ground the screen room to the control room ground to prevent flashover from the primary to secondary winding. Tests, which are not included in the paper, were also made for this same condition, except that the motor-generator power supply was used. These oscillograms show freedom from the disturbances obtained with the isolation transformer. Hence, it may be concluded that the disturbances shown in Fig. 17 are introduced through the isolation transformer and are not due to the inner and outer screen rooms and the CRO cabinet being connected to the control room ground.

Analysis of Nonresonant Charging Circuit for Magnetic Pulse Generator

N. L. WEINBERG
NONMEMBER AIEE

THE USE of saturable reactors to produce narrow, high-energy pulses is becoming increasingly important in radar and similar applications where rugged devices of high reliability are required. General principles of operation of these magnetic pulse generators are discussed thoroughly in the original paper on this subject by W. S. Melville.[1] Subsequently, other articles have been published dealing with design criteria, core volume derivations, and other additions to the theory of operation and state of the art.[2–5] These works, however, concern themselves mainly with resonant charging circuits. The intent of this paper is to derive the fundamental equations for nonresonant operation of the input charging circuit and to compare performance as a function of the ratio of charging circuit frequency to line frequency. Piecewise linear analysis is the method employed to obtain the solutions of this nonlinear circuit.

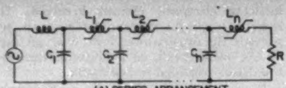
The two most popular basic configurations are illustrated in Fig. 1 (the series line in *A*, the shunt circuit in *B*). Both circuits have the same type of output and the analysis contained herein is valid for both circuits in the ideal case. In the remainder of this paper, however, the discussion will be restricted to the series line which is the most common application of this type of device. The two modes of operation are termed uni-cycle and bi-cycle, with uni-cycle defined as one unidirectional pulse per cycle of input frequency and bi-cycle defined as one positive and one negative pulse per input cycle. Uni-cycle is accomplished physically by biasing the first reactor so that it can saturate only in one direction.

A brief description of circuit operation includes first the line charging unit, L and C_1. In the resonant case the voltage across the capacitor (which equals the voltage across L_1 in the ideal case) has the waveform illustrated in Fig. 2 where both uni-cycle and bi-cycle modes are compared with the source voltage. When the voltage on capacitor C_1 reaches its peak, reactor L_1 saturates and C_1 discharges through the saturated inductance of L_1 charging C_2. When the voltage on C_2 reaches its peak, L_2 saturates and C_2 discharges into the succeeding capacitor C_3. This process continues until the final capacitor C_n discharges through the saturated inductance of the final reactor L_n into the load R. The reactors are designed to saturate when the peak voltage is reached, and the discharge frequency of each stage is determined by the saturated inductance of the reactor and the capacity of the discharge loop. In the ideal pulser as well as in many practical models, the capacitors of each stage (when referred to the same impedance level where interstage transformers are used) are equal. Thus the discharge frequency of each stage is solely a function of the reactor saturated inductance. The saturated inductance of the first reactor is very much smaller than the charging inductance L so that the discharge fre-

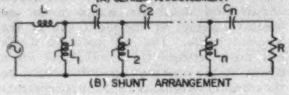
---

Paper 57-963, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted June 5, 1957; made available for printing October 23, 1957.

N. L. Weinberg is with the Westinghouse Electric Corporation, Baltimore, Md.



(A) SERIES ARRANGEMENT



(B) SHUNT ARRANGEMENT

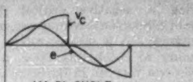
Fig. 1. Basic pulser circuits

quency of the first stage is greater than line frequency. The second capacitor, therefore, charges up faster than the first capacitor, resulting in a second-stage reactor with reduced turns-core area product and, consequently, reduced saturated inductance. The discharge frequency is thus increased over the first stage and pulse sharpening has been obtained. This process of pulse sharpening is continued until a sufficiently narrow pulse, as determined by load specifications, has been obtained in the final stage. Although indicated simply in the diagrams, for radar applications C_n could represent a pulse-forming network and R a pulse transformer driving a magnetron.

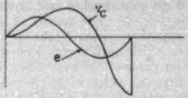
Reduction of size and weight of magnetic pulse generators (or pulsers as they are sometimes called) by core volume reduction is one of the major objectives in this field today. Increasing the pulse sharpening per stage cannot be achieved without using excessively large cores and for specifications on any given pulser, the optimum number of stages is easily determined. A reduction in the volume of the first stage without altering the pulse compression per stage would result obviously in a smaller pulser. Since the volume of the first-stage reactor is proportional to the volt-second integral (for given output pulse from first stage) of the waveshape on C_1, reduction in charging time keeping the peak voltage constant would result in a smaller reactor. By utilizing nonresonant charging, this can be accomplished and in the analyses that follow, the properties of the circuit relevant to pulser design will be derived for the ideal general case for both unicycle and bi-cycle modes.

Nomenclature

A = effective core cross-sectional area, square centimeters
ΔB = "knee-to-knee" flux density, gauss
C = capacitance of each stage, farads
e = instantaneous line voltage, volts
E = peak line voltage, volts
f = line frequency, cycles per second



(A) BI-CYCLE



(B) UNI-CYCLE

Fig. 2. v_c and e for resonant charging

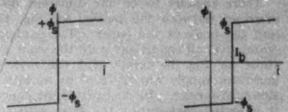
i = line current including both i_1 and i_2, amperes
i_1 = line current for reactor in saturation, amperes
i_2 = line current during capacitor charge, amperes
i_c = capacitor current, amperes
I_b = bias current referred to main winding, amperes
I_{bn} = normalized bias current, amperes
I_f = fundamental component of line current, rms, amperes
I_{rms} = root-mean-square value of line current, amperes
K = reactor turns-core area design factor
L = inductance of charging choke, henrys
L_{sk} = saturated inductance of kth reactor, henrys
$n = (w_0/w)$ = ratio of charging circuit frequency to line frequency
N = number of turns on reactor
P = pulser input power, watts
q = charge on C_1, coulombs
r = pulser repetition rate
t_0 = time at which reactor unsaturates
t_1 = zero crossover for uni-cycle capacitor voltage
v_c = instantaneous voltage on C_1, volts
V_c = peak voltage on C_1
ϕ = phase angle between line voltage and reactor saturation time
w = radian line frequency
$w_0 = (1/LC)^{1/2}$ = radian frequency of input charging circuit

Analysis of Bi-Cycle Mode

In the circuit simplifications that follow, ideal flux-current characteristics as shown in Fig. 3(A) will be assumed for all reactors. To facilitate discussion of the nonresonant case, typical waveforms of source voltage, line current, and capacitor voltage which could be obtained experimentally or by deduction are illustrated in Fig. 4(A). Zero time is chosen as the instant the reactor saturates. Since the phase angle between source voltage and firing point is a function of nonresonance, ϕ must be included in the analysis. All equations will be derived for the polarities illustrated and will be valid for $0 \leq wt \leq \pi$.

Immediately following discharge of capacitor C_1 at $t=0$, the reactors remain in saturation in the ideal case. Although this is not strictly true in actual modulators (since it takes a finite time for the pulse to travel down the chain until it is discharged into the load), the time is very small compared to the period of the input sinusoid. Therefore, the resulting error is negligible. The saturated inductances of the reactors are negligible impedances at line frequency as compared to the capacitor impedance so that during saturation, line current will flow through the reactors rather than the capacitors. In the final stage, load resistance is also negligibly small in comparison to capacitor impedance at line frequency so that the previous assumption is valid for the final stage as well. Since $L \gg L_{S1} \gg L_{S2} \dots \gg L_{SN}$, the impedance of this series chain of inductances can be represented by the charging inductance alone.

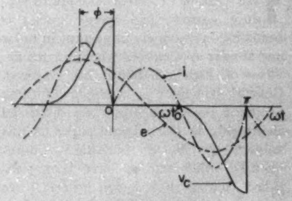
During unsaturation, the reactors are open circuits, since ideal characteristics are assumed, and the circuit becomes merely L and C in series. This assumption of zero exciting current is valid since this current is extremely small compared to circuit currents for the square-loop core materials employed in these pulsers.



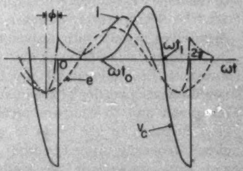
(A) IDEAL FLUX-CURRENT CURVE

(B) IDEAL FLUX-CURRENT CURVE FOR MAIN WINDING CURRENT UNI-CYCLE

Fig. 3. Magnetization curves



(A) BI-CYCLE



(B) UNI-CYCLE

Fig. 4. Waveforms

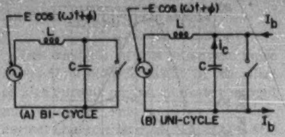


Fig. 5. Equivalent circuits

The equivalent circuit can thus be represented by the configuration shown in Fig. 5(A) with the switch symbolically indicating the two operating regions of the reactor.

Boundary conditions can be determined easily by deduction and confirmed by experiment. Optimum circuit operation requires that the reactor saturate when V_c has been reached. Thus $dv_c(0)/dt=0$ and since current is continuous through L (due to the requirement of continuous flux linkages),

$$i(0)=0 \tag{1}$$

Since the circuit must be periodic at line frequency,

$$i\left(\frac{\pi}{w}\right)=0 \tag{2}$$

At $t=t_0$, the reactor unsaturates and observation of the flux-current curve illustrates that this must occur when $i=0$. Thus,

$$i(t_0)=0 \tag{3}$$

The final boundary condition required is the voltage across the capacitor at $t=t_0$. At the instant the switch opens, charge on the capacitor is zero, and therefore

$$v_c(t_0)=0 \tag{4}$$

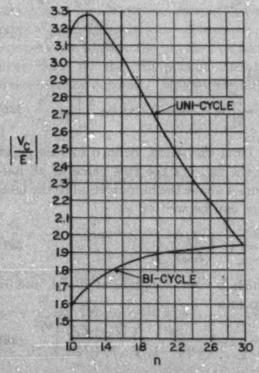


Fig. 6. Capacitor voltage step-up versus nonresonant frequency ratio

The original physical circuit has been formulated as a mathematical problem and its solutions can now be obtained. In the analysis that follows, it is assumed that $n \neq 1$ (the resonant case is not valid in the derivation) and since circuit frequencies greater than line frequency are being considered, $n>1$.

During saturation of the reactor (switch closed in Fig. 5) the circuit equation is

$$L\frac{di_1}{dt}=E\cos(wt+\phi) \tag{5}$$

Separating variables, integrating, and using equation 1,

$$i_1=\frac{E}{wL}[\sin(wt+\phi)-\sin\phi] \tag{6}$$

for $0 \leq t \leq t_0$

An equation relating t_0 and ϕ can be obtained by utilizing equation 3 resulting in

$$\sin(wt_0+\phi)=\sin\phi \tag{7}$$

To obtain the solution valid in the interval $(0, \pi/w)$, the supplement of ϕ must be employed in equation 7. Equating the appropriate arguments,

$$wt_0=\pi-2\phi \tag{8}$$

For the period of capacitor charge, the integrodifferential equation is

$$L\frac{di_2}{dt}+\frac{\int i_2 dt}{C}=E\cos(wt+\phi) \tag{9}$$

Rewriting this equation in terms of charge and using $w_0^2LC=1$,

$$\frac{d^2q}{dt^2}+w_0^2q=\frac{E}{L}\cos(wt+\phi) \tag{10}$$

This is a linear second-order ordinary differential equation whose solution can be readily obtained by any of several elementary methods. If Laplace transforms are employed, care must be exercised in defining the functions so that initial conditions at $t=t_0$ can be used.

Using equations 3 and 4, the solution of equation 9 is determined as

$$q(t)=\frac{n^2CE}{(n^2-1)}\Big[\cos\phi\cos n(wt+2\phi-\pi)+ \frac{1}{n}\sin\phi\sin n(wt+2\phi-\pi)+ \cos(wt+\phi)\Big] \tag{11}$$

Differentiating both sides of equation 11,

$$i_2(t)=\frac{n^2wCE}{(n^2-1)}[\sin\phi\cos n(wt+2\phi-\pi)- n\cos\phi\sin n(wt+2\phi-\pi)- \sin(wt+\phi)] \tag{12}$$

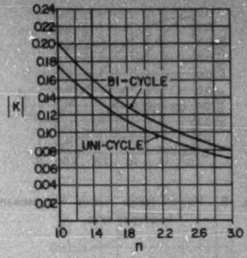


Fig. 7. Reactor turns-area design factor versus nonresonant frequency ratio

Utilizing equation 2, the following trigonometric relation is obtained:

$$\sin\phi\cos 2\phi n-n\cos\phi\sin 2\phi n- \sin(\pi+\phi)=0 \tag{13}$$

The expression in equation 13 can be factored leading to

$$\phi=\pi/2n \tag{14}$$

Substituting the value of ϕ from equation 14 into equations 11 and 12, the solutions for capacitor voltage and line current are obtained during the period of capacitor charge.

$$v_c(t)^{\dagger}=\frac{n^2E}{(n^2-1)}\Big[\cos\left(wt+\frac{\pi}{2n}\right)- \cos\frac{\pi}{2n}\cos n(wt-\pi)- \frac{1}{n}\sin\frac{\pi}{2n}\sin n(wt-\pi)\Big] \tag{15}$$

$$i_2(t)=\frac{n^2wCE}{(n^2-1)}\Big[n\cos\frac{\pi}{2n}\sin n(wt-\pi)- \sin\frac{\pi}{2n}\cos n(wt-\pi)- \sin\left(wt+\frac{\pi}{2n}\right)\Big] \tag{16}$$

$wt_0 \leq wt \leq \pi$

Inserting the value of ϕ obtained from equation 14 into equation 6, the solution for line current during reactor saturation is

$$i_1(t)=n^2wCE\left[\sin\left(wt+\frac{\pi}{2n}\right)-\sin\frac{\pi}{2n}\right] \tag{17}$$

Use was made of the relation $1/wL=n^2wC$ in deriving equation 17. Using equation 14, t_0, the point at which the reactor unsaturates, can be expressed as a function of n, namely,

$$wt_0=\frac{(n-1)\pi}{n} \tag{18}$$

Peak capacitor voltage, V_c, can be obtained easily from equation 15; for since

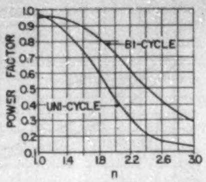


Fig. 8. Power factor versus nonresonant frequency ratio

$$V_c = v_c\left(\frac{\pi}{w}\right)$$

it follows that

$$V_c = \frac{-2n^2E}{n^2-1}\cos\frac{\pi}{2n} \tag{19}$$

with the negative sign the result of choosing polarities as in Fig. 4(A). A curve of $|V_c/E|$ as a function of nonresonance ratio is illustrated in Fig. 6.

One of the most important considerations for the design and evaluation of the first-stage reactor is the turns-area product required for saturation to occur at the proper point of the voltage waveform across the reactor.

From Faraday's law, one can derive

$$NA = \frac{-(10)^8}{\Delta B}\int v_c(t)\,dt \tag{20}$$

For the waveforms under examination, this integral is

$$\int_0^{\frac{\pi}{w}} v_c(t)\,dt$$

which in view of equation 15 becomes

$$\int_{t_0}^{\frac{\pi}{w}} v_c(t)\,dt = \frac{-2E}{w}\sin\frac{\pi}{2n} \tag{21}$$

Substituting this value in equation 20,

$$NA = \frac{2(10)^8E}{w\Delta B}\sin\frac{\pi}{2n} \tag{22}$$

In any practical pulser, circuit losses will reduce the theoretical charging circuit step-up ratio (V_c/E) indicated by equation 19 so that writing equation 22 in terms of V_c would be more useful. Making the appropriate substitution from equation 19, equation 22 becomes

$$NA = \frac{-(n^2-1)\,V_c(10)^8\tan\dfrac{\pi}{2n}}{n^2w\Delta B}$$

Using the familiar form for this quantity,

$$NA = K\,\frac{V_c(10)^8}{\Delta Bf} \tag{23}$$

where

$$K = -\frac{(n^2-1)}{2\pi n^2}\tan\frac{\pi}{2n} \tag{24}$$

A plot of $|K|$ as a function of nonresonance ratio is shown in Fig. 7.

Another important consideration in these high-powered pulsers is line power factor. Due to the large power drain from the source, poor power factor would be detrimental in applications where generator size is critical, e.g., in aircraft and missiles. In order to calculate power factor, rms current is needed and defined in the usual manner,

$$I_{\text{rms}} = \left[\frac{w}{\pi}\left(\int_0^{t_0} i_1^2dt + \int_{t_0}^{\pi/w} i_2^2dt\right)\right]^{1/2} \tag{25}$$

Evaluating the integrals in equation 25 using equations 16, 17, and 18,

$$I_{\text{rms}} = nwCE[F_1+F_2]^{1/2} \tag{26}$$

where

$$F_1 = \frac{n}{2}\left[(n-1)\left(2-\cos\frac{\pi}{\pi}\right) - \frac{3n}{\pi}\sin\frac{\pi}{n}\right] \tag{27}$$

and

$$F_2 = \frac{n}{4(n^2-1)^2}\left[(n^2+3)+(n^2-1)\cos\frac{\pi}{n} - \frac{2n(n^2+3)}{\pi(n^2-1)}\sin\frac{\pi}{n}\right] \tag{28}$$

Since the lossless circuit is being considered, power input equals the power discharged by capacitor C_1 which is

$$\frac{1}{2}CV_c^2r \tag{29}$$

In the bi-cycle case, $r=2f$ and therefore, $P=CV_c^2f$. Substituting for V_c,

$$P = \frac{2wCn^4E^2}{\pi(n^2-1)^2}\cos^2\frac{\pi}{2n} \tag{30}$$

Using equation 26, input volt-amperes is

$$nwCE^2\left[\frac{F_1+F_2}{2}\right]^{1/2} \tag{31}$$

From equations 30 and 31,

$$\text{Power factor} = \frac{2n^3\cos^2\dfrac{\pi}{2n}}{\pi(n^2-1)^2}\left[\frac{2}{F_1+F_2}\right]^{1/2} \tag{32}$$

where F_1 and F_2 are defined by equations 27 and 28. A graph of power factor (pf) versus n is shown in Fig. 8.

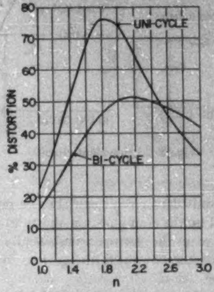


Fig. 9. Line-current distortion versus nonresonant frequency ratio

The final derivation to be made in the bi-cycle mode is the calculation of line-current distortion. Due to generator characteristics in certain systems, high harmonic content in the pulser input current could distort the voltage waveform of the source disturbing the operation of the other equipment. In order to calculate distortion, the rms amplitude of the fundamental is required and this can be easily obtained from the coefficients of the fundamental components of the Fourier series. Defining these quantities in the usual way and using the property for both i_1 and i_2 that

$$i(t) = -i\left(t-\frac{\pi}{w}\right)$$

$$I_f = \frac{w\sqrt{2}}{\pi}\left[\left(\int_0^{\pi/w} i(t)\cos wtdt\right)^2 + \left(\int_0^{\pi/w} i(t)\sin wtdt\right)^2\right]^{1/2} \tag{33}$$

Performing the required integrations, one obtains

$$I_f = \frac{n^2wCE}{\pi(n^2-1)}[2(F_3^2+F_4^2)]^{1/2} \tag{34}$$

where

$$F_3 = \frac{1}{2}\left[\pi(n^2-n-1) - n^2\sin\frac{\pi}{n}\right]\sin\frac{\pi}{n} + \frac{2n^2}{n^2-1}\cos\frac{\pi}{2n} \tag{35}$$

and

$$F_4 = \frac{1}{2}\left[\pi(n^2-n-1) - n^2\sin\frac{\pi}{n}\right]\cos\frac{\pi}{n} \tag{36}$$

Total distortion, defined in this paper as the ratio of the rms value of all harmonics above the fundamental to the total rms, can be written as

$$D = \left[1-\left(\frac{I_f}{I_{\text{rms}}}\right)^2\right]^{1/2} \tag{37}$$

which in view of equations 26, 34, and 37 becomes

$$D = \left[1-\left(\frac{2n^2}{\pi^2(n^2-1)^2}\right)\frac{(F_3^2+F_4^2)}{(F_1+F_2)}\right]^{1/2} \tag{38}$$

A plot of D versus n is illustrated in Fig. 9.

Analysis of Uni-Cycle Mode

The analysis of the uni-cycle mode differs from bi-cycle chiefly in the effect of bias current on the equivalent circuit and boundary values. The ideal flux-current characteristic of Fig. 3(A) is still applicable; however, current must now be interpreted as the total reactor current. Assuming a 1-to-1 turns ratio between main and bias windings, the flux-current characteristic for the current in the main winding is as illustrated in Fig. 3(B) with I_b representing the d-c bias current. Thus, the major change from bi-cycle operation is the current drawn by the reactor during unsaturation. In the equivalent circuit, this additional branch current must be taken into account as shown in Fig. 5(B). Polarities for the analysis are selected as in the waveforms of Fig. 4(B).

Boundary values are deduced as in the bi-cycle case and require revision only with respect to the additional branch current, I_b. At $t=0$, $dv_c/dt=0$ so that

$$i_c(0)=0 \tag{39}$$

Since $i=i_c+I_b$,

$$i(0)=I_b \tag{40}$$

The other conditions are

$$i(t_0)=I_b \tag{41}$$

$$i\left(\frac{2\pi}{w}\right)=I_b \tag{42}$$

and

$$v_c(t_0)=0 \tag{43}$$

In the bi-cycle case, waveforms are symmetrical over the entire cycle which results in a zero d-c component. In the uni-cycle mode, however, current waveforms are not symmetrical. Therefore, it is imperative to state the physical requirement that no d-c component can be present in line current. Mathematically,

$$\int_0^{\frac{2\pi}{w}} i(t)dt=0 \tag{44}$$

The circuit equation during saturation is

$$L\frac{di_1}{dt}=-E\cos(wt+\phi) \tag{45}$$

and the solution using equation 40 is

$$i_1=-\frac{E}{wL}[\sin(wt+\phi)-\sin\phi]+I_b \tag{46}$$

for $0\leq t\leq t_0$

Using equation 41, one obtains

$$wt_0=\pi-2\phi \tag{47}$$

as the solution for t_0 in the interval $(0, 2\pi/w)$.

During capacitor charge, the equation becomes

$$L\frac{d}{dt}(ic+I_b)+\frac{\int i_c dt}{C}=-E\cos(wt+\phi) \tag{48}$$

Since I_b is constant, this reduces to

$$L\frac{di_c}{dt}+\frac{\int i_c dt}{C}=-E\cos(wt+\phi) \tag{49}$$

This equation, other than for polarity of driving voltage, is identical to the bi-cycle mode equation, and the solution can be written directly. Since $i_2(t)=i_c+I_b$, the solution for line current is

$$i_2(t)=I_b+\frac{n^2wCE}{(n^2-1)}[n\cos\phi\sin n\times(wt+2\phi-\pi)-\sin\phi\cos n\times(wt+2\phi-\pi)+\sin(wt+\phi)] \tag{50}$$

and capacitor voltage is

$$v_c(t)=-\frac{n^2E}{(n^2-1)}\left[\cos\phi\cos n\times(wt+2\phi-\pi)+\frac{1}{n}\sin\phi\sin n\times(wt+2\phi-\pi)+\cos(wt+\phi)\right] \tag{51}$$

Using equation 42 in equation 50, one obtains

$$n\cos\phi\sin n(\pi+2\phi)-\sin\phi\times\cos n(\pi+2\phi)+\sin\phi=0 \tag{52}$$

This can be factored leading to the solution.

$$\phi=\frac{\pi}{n}-\frac{\pi}{2} \tag{53}$$

Substituting this value of ϕ in equations 50 and 51,

$$i_2(t)=I_b+\frac{n^2wCE}{(n^2-1)}\left[n\sin\frac{\pi}{n}\sin n(wt-2\pi)+\cos\frac{\pi}{n}\cos n(wt-2\pi)-\cos\left(wt+\frac{\pi}{n}\right)\right] \tag{54}$$

and

$$v_c(t)=-\frac{n^2E}{(n^2-1)}\left[\sin\frac{\pi}{n}\cos n(wt-2\pi)-\frac{1}{n}\cos\frac{\pi}{n}\sin n(wt-2\pi)+\sin\left(wt+\frac{\pi}{n}\right)\right] \tag{55}$$

Substituting this value of ϕ from equation 53 into equation 46, and using $1/wL=n^2wC$,

$$i_1(t)=n^2wCE\left[\cos\left(wt+\frac{\pi}{n}\right)-\cos\frac{\pi}{n}\right]+I_b \tag{56}$$

and the point of unsaturation becomes

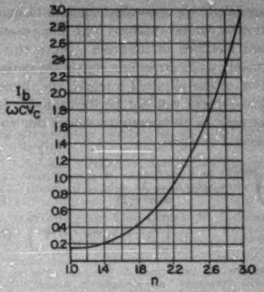


Fig. 10. Normalized bias current versus nonresonant frequency ratio

$$wt_0=\frac{2\pi(n-1)}{n} \tag{57}$$

Using equation 55,

$$V_c=v_c\left(\frac{2\pi}{w}\right)=\frac{-2n^2E\sin\frac{\pi}{n}}{n^2-1} \tag{58}$$

and a plot of $|V_c/E|$ versus n is illustrated in Fig. 6.

To determine I_b, equation 44 will be utilized. Therefore,

$$0=\int_0^{\frac{2\pi}{w}} i(t)dt=\int_0^{t_0} i_1(t)dt+\int_{t_0}^{\frac{2\pi}{w}} i_2(t)dt \tag{59}$$

Evaluating the integrals and solving,

$$I_b=n^2wCE\left[\frac{n^2}{\pi(n^2-1)}\sin\frac{\pi}{n}+\frac{(n-1)}{n}\cos\frac{\pi}{n}\right] \tag{60}$$

As in the case for NA, bias current in terms of V_c rather than E would be more useful and informative. Making the appropriate substitution,

$$I_b=-\frac{wCV_c}{2}\left[\frac{n^2}{\pi}+\frac{(n^2-1)(n-1)}{n}\cot\frac{\pi}{n}\right] \tag{61}$$

A curve of normalized bias current, I_b/wCV_c versus n (disregarding polarity) is illustrated in Fig. 10.

Computation of the volt-second integral for evaluating NA is more complicated for uni-cycle than bi-cycle due to the reversal of polarity of $v_c(t)$. A necessary preliminary, therefore, is the determination of the zero crossover, t_1. Since $v_c(t_1)=0$,

$$\sin\frac{\pi}{n}\cos n(wt_1-2\pi)-\frac{1}{n}\cos\frac{\pi}{n}\sin n\times(wt_1-2\pi)+\sin\left(wt_1+\frac{\pi}{n}\right)=0 \tag{62}$$

Unfortunately, equation 62 cannot be solved explicitly for t_1. However, for any specified n, a numerical solution of equation 62 can be obtained for t_1 using standard methods of iteration. Since this is merely an intermediate step in the derivation of the required integral, the numerical process will not be illustrated. Returning to the computation of the volt-second integral,

$$\int_{t_1}^{2\pi/w} v_c(t)\,dt = \frac{E}{w(n^2-1)} \times \left[(n^2-1)\cos\frac{\pi}{n} + F_5\right] \qquad (63)$$

where

$$F_5 = n \sin\frac{\pi}{n}\sin n(wt_1 - 2\pi) + \cos\frac{\pi}{n}\cos n(wt_1-2\pi) - n^2\cos\left(wt_1 + \frac{\pi}{n}\right) \qquad (64)$$

Utilizing equation 62, F_5 can be evaluated numerically as a function of n.

Inserting V_1 from equation 58 into equation 63,

$$\int_{t_1}^{2\pi/w} v_c(t)\,dt = -\frac{V_c}{2n^2 w \sin\frac{\pi}{n}} \times \left[(n^2-1)\cos\frac{\pi}{n} + F_5\right] \qquad (65)$$

Using the same form for NA as in equation 23,

$$K = -\frac{(n^2-1)\cos\frac{\pi}{n} + F_5}{4\pi n^2 \sin\frac{\pi}{n}} \qquad (66)$$

A plot of $|K|$ for the uni-cycle mode of operation is illustrated in Fig. 7. Calculating I_{rms} for computation of power factor,

$$I_{rms} = n^2 wCE[F_6 + F_7]^{1/2} \qquad (67)$$

where

$$F_6 = \frac{(n-1)}{n}\left[\frac{1}{2} + \left(I_{bn} - \cos\frac{\pi}{n}\right)^2\right] + \frac{1}{2\pi}\left[3\cos\frac{\pi}{n} - 4I_{bn}\right]\sin\frac{\pi}{n} \qquad (68)$$

$$F_7 = \frac{I_{bn}^2}{n} + \frac{n^2+3}{4n(n^2-1)^2} - \frac{1}{4n(n^2-1)} \times \cos\frac{2\pi}{n} - \frac{2I_{bn}}{\pi(n^2-1)}\sin\frac{\pi}{n} + \frac{(n^2+3)}{4\pi(n^2-1)^3}\sin\frac{2\pi}{n} \qquad (69)$$

and

$$I_{bn} = \frac{I_b}{n^2 wCE} \qquad (70)$$

From equation 67,

Input volt-amperes

$$= n^2 wCE^2\left[\frac{F_6+F_7}{2}\right]^{1/2} \qquad (71)$$

Since in uni-cycle there is only one pulse per cycle,

$$P = \frac{1}{2} CV_c^2 f \qquad (72)$$

Using equation 58,

$$P = \frac{n^4 wCE^2}{\pi(n^2-1)^2}\sin^2\frac{\pi}{n} \qquad (73)$$

Utilizing equations 71 and 73,

$$\text{Power factor} = \frac{n^2}{\pi(n^2-1)^2} \times \left[\frac{2}{F_6+F_7}\right]^{1/2}\sin^2\frac{\pi}{n} \qquad (74)$$

The curve of pf is illustrated in Fig. 8.

To calculate distortion, the fundamental rms amplitude of line current is required. Deriving this quantity, one obtains

$$I_f = n^2 wCE\left[\frac{F_8^2+F_9^2}{2}\right]^{1/2} \qquad (75)$$

where

$$F_8 = \frac{1}{(n^2-1)}\left[(n^2-n-1)\cos\frac{\pi}{n} + \frac{n^2}{2\pi}\cos\frac{\pi}{n}\sin\frac{2\pi}{n} - \frac{2n^2}{\pi(n^2-1)} \times \sin\frac{\pi}{n}\right] \qquad (76)$$

and

$$F_9 = \frac{1}{(n^2-1)}\left[(n^2-n-1)\sin\frac{\pi}{n} + \frac{n^2}{2\pi}\sin\frac{\pi}{n}\sin\frac{2\pi}{n}\right] \qquad (77)$$

Calculating total distortion as in bi-cycle and using equations 67 and 75,

$$D = \left[1 - \frac{F_8^2+F_9^2}{2(F_6+F_7)}\right]^{1/2} \qquad (78)$$

D versus n is shown in Fig. 9.

Discussion

Examination of Figs. 6–10 reveals many interesting aspects of nonresonant charging. It is apparent from Fig. 7, that a faster charging rate for the input stage results in a considerable reduction of first-stage reactor NA and consequently of core volume. This characteristic applies for both modes of operation.

Fig. 6 indicates that capacitor voltage step-up increases with charging circuit frequency in the bi-cycle case up to a maximum step-up of 2.0. For the uni-cycle mode, however, after the maximum is reached at a ratio of 1.2, voltage step-up is reduced with increasing frequency. In evaluating performance, this characteristic is actually unimportant since, in most pulsers, an input transformer is required to raise the voltage to the desired level and the transformer turns ratio can be altered to take the actual step-up voltage into account. In practical pulsers the values in Fig. 6 will have to be revised to take into account actual circuit losses, such as those due to the charging choke and the core loss in the first stage reactor.

The pf curves of Fig. 8 illustrate another difference between the two modes of operation. In bi-cycle, pf actually rises as n is increased from resonance and remains above 90% until $n=1.55$ is exceeded. For frequency ratios to 1.8, pf's greater than 80% can be obtained. In uni-cycle, however, pf drops off sharply as n is increased above resonance. To obtain a pf of 90%, n must be limited to less than 1.3, and for a pf of 80%, n cannot exceed 1.5. On the basis of pf alone, the n which could be utilized in the uni-cycle mode is limited to a fairly low value.

The nature of line-current distortion, as illustrated in Fig. 9, is quite evident. In applications where the harmonic content of the current drawn from the generator is critical, nonresonance could not be employed. In many applications, however, distortion has not been critical and interference with other equipment using the same generator has been negligible. As one would expect, the nonsymmetrical waveforms of uni-cycle operation result in higher distortion than for bi-cycle. In actual pulsers, the cores do not saturate so sharply as shown in the ideal magnetization curves utilized in the derivation; consequently, distortion is usually less than that predicted from the analysis.

The bias current curve of Fig. 10 offers another reason why higher values of n are impractical in the uni-cycle mode. Bias current rises gradually with nonresonance, and for $n=1.5$ an increase of approximately 50% is obtained. For greater value of n, bias increases sharply and becomes excessive.

The principal advantage of nonresonant charging circuits is the reduction in core volume accomplished by faster charging rates. For bi-cycle operation, $n=1.5$ results in a reduction of 25% of NA and in a pf greater than 90%. In the unicycle case, the reduction in NA is approximately the same, but due to lower pf, higher distortion, and increased bias, n must be limited to values less than 1.3. In practice, the choice of charging circuit frequency depends on a compromise between reduced core volume and increased demand on the power supply.

Performance at resonance can be obtained either by using the basic equations of the resonant circuit or by taking the limit as n approaches unity in the equations of this paper. Although the derivation is not valid for $n=1$, the solution for resonance exists as a limiting value. In the case of the more involved equations, it is simpler to use the resonant circuit equations. The limiting process, however, is nevertheless advisable to provide a check on the validity of the analysis. The ideal magnetization curve is a close approximation for equipment of this type and good agreement has been obtained with experimental data. It should be understood that the curves presented here are not intended as accurate design criteria but merely indicate general performance and trends. The results of an ideal analysis are sufficiently close so that equipment based on it should require only minor modifications.

Conclusions

A nonresonant charging circuit for magnetic pulse generators has been analyzed by employing piecewise linear methods. The most pertinent design considerations have been summarized graphically for both the uni-cycle and bicycle modes of operation. It has been shown that the advantages gained in core volume reduction must be weighed against increased demands on the power supply.

References

1. The Use of Saturable Reactors as Discharge Devices for Pulse Generators, **W. S. Melville.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 98, pt. III, 1951, pp. 185–207.

2. Economic Design of Saturating Reactor Magnetic Pulsers, **R. A. Mathias, E. M. Williams.** *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 169–71.

3. Magnetic Pulse Modulators, **K. J. Busch, A. D. Hasley, C. Neitzert.** *Bell System Technical Journal*, New York, N. Y., Sept. 1955, pp. 943–93.

4. Core Volume Derivation for Magnetic Pulse Modulators, **J. E. Sunderlin.** *Proceedings*, IRE National Conference on Aeronautic Electronics, Dayton, Ohio, 1956, pp. 11–16.

5. Magnetic Pulse Generator Practical Design Limitations, **M. F. Thompson, R. R. Trautwein, E. R. Ingersoll.** *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 786–92

The Definition and Measurement of the Time Constant and Response Time of Thermal Converters

FRANCIS L. HERMACH
MEMBER AIEE

Synopsis: The dynamic characteristics of a thermal converter for measuring alternating current or voltage are of some importance in a number of applications. The time constant and response time are the characteristics most often required. The change of emf (electromotive force) of a thermal converter after a step change in heater current is only approximately exponential, so that the definition as well as methods of measuring these quantities must be specified. This investigation of methods of measuring these quantities was undertaken at the request of the AIEE Committee on Indicating and Integrating Instruments.

A THERMAL CONVERTER (often called a thermoelement or thermocouple) consists of a thermojunction in thermal contact with an electric heater, so that the output emf serves as a measure of the heater current.[1] Thermal converters are finding increasing use as primary elements for measuring alternating current and voltage, because of their d-c output and wide frequency range, and as a-c–d-c transfer standards, because the transfer may be made very accurately with properly constructed converters. They are also valuable as detectors for noise studies (with suitable amplifiers), because they integrate to give the rms (root-mean-square) value of a signal. In each application the dynamic characteristics are of some importance. These characteristics are usually evaluated in terms of either a time constant of the change of emf after a step change in heater current, or a response time for essentially all of the change to take place.

The time constant t_c of a simple exponential function of an independent variable t is the value of t required for $100(\epsilon-1)/\epsilon$ per cent (63.2%) of the total change in the dependent variable, where ϵ is the base of the natural logarithms. However, for most converters the change of emf is only approximately exponential. Thus a definition must precede the selection of a method for making the measurements. One definition, which we shall call t_a, is the time in seconds required for 63% of the change to take place. Another which has been proposed is given by the expression $t_b=(1/E)\int_0^\infty e\,dt$, where e is the difference between the final emf and the actual emf at the time t, and E is the difference between the final and initial emf's, as shown in Fig. 1. Both of these reduce to t_c for a simple exponential function.

A dynamic characteristic of interest in ordinary instrument applications is the response time. This is defined for an indicating instrument[1] as the time required for the pointer of the instrument to come to apparent rest after an abrupt change has occurred in the measured quantity. The pointer is considered to have come to apparent rest when it remains within a range on either side of its final position equal to 1/2 the accuracy rating of the instrument.

When a thermal converter is used as a primary element in an instrument, its response time may be more important than its time constant. The usual thermocouple instruments are in about 1% accuracy class and the scale of a thermocouple instrument is nearly square law, so that $dD/D=2dI/I$, where D is the instrument deflection and I is the heater current. Since the deflection is proportional to the emf of the converter it follows that the time required for 90% of the change in emf to take place could appropriately but somewhat arbitrarily be considered as the response time. However, at least one manufacturer has for a

---

Paper 58-11, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted March 25, 1957; made available for printing October 23, 1957.

Francis L. Hermach is with the National Bureau of Standards, Washington, D. C.

The author acknowledges with thanks the help of John H. Miller, who first indicated the scope of the problem, furnished special long-time-constant converters, and made available the results of measurements at the Weston Electric Instrument Corporation; and the help of James E. Griffin, who checked the equations in the appendixes.

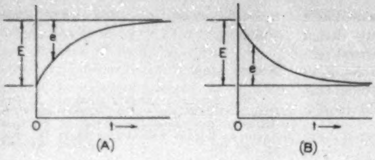


Fig. 1. Change of emf after step change in current

A—Increasing emf
B—Decreasing emf

number of years used the 90%-response time in specifying the dynamic characteristics of his thermal converters, perhaps because this is easier to measure. There is thus a need for precise definitions and experimental studies of actual converters.

Measurements

Several methods of measuring the dynamic characteristics of thermal converters were considered, from the viewpoints of practicality, accuracy, and the commercial availability of apparatus. The time constants and response times of a group of several types of thermal converters were measured by two of the most promising methods, with every reasonable effort to obtain an accuracy of 2% in the measurement. These converters had time constants ranging from 0.1 to 10 seconds.

GALVANOMETER METHOD

The first of these two methods consisted simply of obtaining and scaling oscillograms of the output emf of each converter as a function of elapsed time after a prescribed step change in heater current. Step changes from zero to the prescribed current and vice versa were used (current switched on and off, respectively). The records were made with a direct-writing electromagnetic oscillograph consisting of a "penmotor" galvanometer with a chart drive of 5, 25, and 125 mm (millimeters) per second and a d-c amplifier which provided about 0.7 of critical damping for the galvanometer. The over-all gain of the amplifier was such that a chart deflection of 40 divisions (40 mm) could easily be obtained at 5 millivolts, with a noise level of about 10 microvolts.

To eliminate errors caused by nonuniformity of the response of the amplifier or the galvanometer and by drifts in the gain of the amplifier, oscillograms were taken, immediately after each measurement, of calibrated direct voltages equal to 37% and 63% of the emf of the converter. To verify the accuracy of the chart drive, oscillograms were periodically taken of amplified seconds pulses from the National Bureau of Standards radio station *WWV*. To reduce timing errors, a second timing pen was energized when the step change in heater current was applied, and the coincidence of the two pens was occasionally carefully checked.

The circuit for making these measurements, with heater current switched either on or off, is shown in Fig. 2. Twisted pair wires were used to minimize inductive loops in both the heater and thermocouple circuits. With switches S_1 and S_4 closed, S_3 open, and S_2, which controlled the timing pen, closed to the left, the heater current I_h was adjusted to the desired value, and the Lindeck-potentiometer current I_p was adjusted for a null balance of the indicating galvanometer G. Switches S_4 and S_5 were then opened, S_3 was closed, the chart drive was energized, and S_1 was first opened and then closed to obtain oscillograms of decreasing and increasing emf. S_5 was then reversed and, with S_1 open, I_p was successively set at 37% and 63% of the value at balance, to calibrate the oscillograph. The time constant t_a and the 90% response time were measured from these calibrated oscillograms.

The 99% response times of a few converters were measured by increasing the gain of the amplifier and offsetting the zero so that a galvanometer deflection of about 30 divisions was obtained for about 10% of the full emf of the converter. The measurements were then repeated along with a calibrating trace differing by about 3% from the full emf.

It is estimated that the galvanometer deflection corresponding to 63% of the change in emf could be determined at best to 0.2 mm on the 40-mm voltage and the chosen time scale. The resultant uncertainties in the measurement of the time constant in seconds are believed to be less than $\pm(2\%+0.4/p)$, where p is the chart speed: 5, 25, or 125 mm/second. Other sources of error are believed to be much less than this for most of the converters tested.

The speed response of the galvanometer system determines, of course, the lower limit of the time constant which may be measured with a given error by the oscillograph. This limit was calculated for an exponential forcing function of time constant t_e applied to a critically damped galvanometer of period t_0. As shown in Appendix I the error should be less than 3% if $t_0/t_e<0.1$, which was true for all but one of the converters tested. While this galvanometer was underdamped and some nonlinearity was introduced in the amplifier by the manufacturer to increase its frequency range for sinusoidal signals, it is believed that this indicates closely enough the error that may be expected from this cause.

FLUXMETER METHOD

A fluxmeter based on the "Inductronic Amplifier"[2] has been used by engineers of the Weston Electrical Instrument Corporation to measure time constants of thermal converters. Although this instrument was not immediately at hand, a commercial, multirange pointer-type fluxmeter was available, and proved to be entirely suitable for measurements of t_b. Such a fluxmeter consists of a permanent-magnet-moving-coil instrument having ligaments of very little restoring torque, rather than spiral springs, to carry the current to the moving coil. Under appropriate conditions the change of deflection of the instrument in a circuit of fixed resistance is $D=K\int e\,dt$ where e is the net emf in the circuit. Thus if the instrument is connected to the output of a thermal converter in such a way that e is the desired change in emf, we have $t_b=(1/E)\int e\,dt=D/EK$. For a converter with a 10-second time constant, the fluxmeter must integrate correctly over a period of about 50 seconds. The conditions which must be met to assure this are considered in Appendix II.

The 5-range, portable fluxmeter available had sensitivities from 0.1 to 10 millivolt-seconds per division in a 15-ohm circuit, with a 100-division mirrored scale. It was used to measure the time constants of the same group of converters, with the circuit shown in Fig. 3. To make a

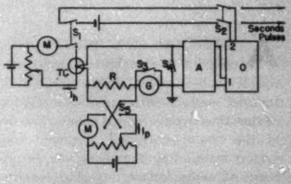


Fig. 2. Circuit for oscillographic measurement

TC—Thermal converter under test
O—Galvanometer oscillograph
A—D-c amplifier
G—Galvanometer
M—Milliammeter

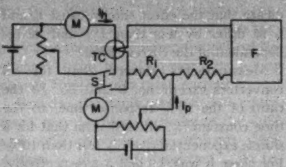


Fig. 3. Circuit for fluxmeter measurement

F—Fluxmeter
TC—Thermal converter
M—Milliammeter

measurement, switch S was closed, the heater current I_h was adjusted to the desired value, I_p was adjusted for a balance as indicated by absence of motion of the pointer of the fluxmeter (on the most sensitive range), and R_2 was set so that the resistance of the circuit as seen by the fluxmeter (including the thermocouple) was 15±0.2 ohms, the design value. The switch was then successively opened and closed and the changes in deflection of the fluxmeter pointer were observed. The constant K was determined at the beginning and end of the entire series of tests by measuring with an electric timer the time required for a given deflection of the pointer, with a constant measured current I_p, with R_2 set as above, and with $I_h=0$. To make sure that the switch blades operated simultaneously, a 4-terminal resistor was substituted for the thermal coverter and the switch was opened and closed several times with I_p set for a balance on the most sensitive range of the instrument. The changes in deflections did not exceed 0.1 division on the scale.

OTHER METHODS

A method of measuring t_b has been developed in which the rise of emf of a converter is opposed to the rise of voltage in an RC circuit, with a ballistic galvanometer as the detector. (Private communication from R. Gilbert to the author.) The circuit diagram is shown in Fig. 4. The resistor R_1 is adjusted to secure a steady state null, and C is then adjusted to produce a transient null when the switch is operated. The null point is defined as the condition of zero integral of voltage at the balance detector. When $R_2 >> R_1$ the R_1C product (in ohms and farads) is the time constant t_b of the converter. The method is not readily adaptable to the full range of time constants of 0.1 to 10 seconds encountered in converters, and it requires a very sensitive ballistic galvanometer, so that the fluxmeter method was used in these tests instead.

The time constant t_a and response time can be measured with a suitable relay which is connected directly to the output of the converter and adjusted to close at the desired millivolt level, to control an electric interval timer. This method, which seems promising, was not tried, because suitable high-speed relays and d-c amplifiers were not readily available.

For some applications the response of a converter to a low-frequency sinusoidal alternating current is of interest. At very low frequencies large-amplitude periodic oscillations of emf occur; the amplitude of these oscillations can be used to estimate the time constant. Calculations show that for converters which are perfectly "square-law," frequencies of about $0.2/t_e$ cycles per second are required to produce an oscillation having a crest value which is 37% of the average value of the output emf. However, the resulting frequency range of 0.02 to 2 c/s is a difficult one in which to work. In addition, nonlinearities in practical converters would probably greatly impair the accuracy of computing t_e from frequency response curves, or vice versa.

Results

The results of measurements of the time constants of representative thermal converters by both methods are given in Table I. The results of measurements of the 99% response time t_r by the oscillographic method are also given. The converters are identified in the table by type and are further described as follows: all converters below 1,000 ma (milliamperes), except no. 7 and no. 8, are of the vacuum type, mounted in an evacuated glass bulb. The 1- and 5-ampere converters have short solid-wire or tubular heaters connected between comparatively large copper blocks. Each bead-type converter has a small electrically insulating bead between the heater and the hot junction of the thermocouple. In the contact type the heater and thermocouple are electrically connected. In the bridge type they are in the same bridge circuit (measurements were made with 60 cycles per second alternating current through the heater). Converter no. 2 is an ultrahigh-frequency type having the heater in line with the lead-in wires. Converter no. 3 is a converter with a heater 1 centimeter long (about twice normal) to obtain good low-frequency performance. Converter no. 4 has an unusually large insulating bead to obtain a long time constant.

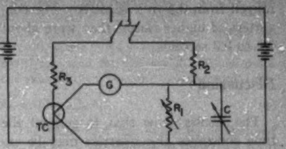


Fig. 4. Circuit for "bridge" method

G—Galvanometer
TC—Thermal converter

The measured time constants of the converters ranged from 0.1 second for a bridge-type converter to over 8 seconds for the special converter having the large insulating bead between the heater and thermocouple. For the usual converters in the milliampere ranges, time constants

Table I. Results of Tests of Thermal Converters

Thermal Converter No.	Rated Current, Ma	Type	I_h, Ma	Time Constant, Seconds					t_r Secs	
				t_a		t_b		t_b/t_a		
				I On	I Off	I On	I Off	I On	I On	$t\ /t_a$
1	5	bead	3	0.86	0.90	0.87	0.94	1.01		
			5	0.80	0.89	0.79	0.90	0.99	3.3	4.1
2	5	bead	3	1.72	1.80	1.71	1.81	1.00		
			5	1.49	1.64	1.44	1.74	0.97		
3	7.5	bead	7.5	2.35	2.62	2.28	2.63	0.97		
4	20	bead	15	8.2	8.4	8.0	8.4	0.98		
			20	7.7	8.5	7.8	8.6	1.01		
5	30	bead	10	1.49	1.48					
			20	1.40	1.47	1.38	1.44	0.99		
			30	1.25	1.38	1.24	1.37	0.99	5.4	4.3
6	100	bead	60	0.63	0.61	0.62	0.63	0.98		
			100	0.60	0.60	0.60	0.61	1.00		
7	100	bridge	100	0.107	0.102	0.107	0.101	1.00		
8	300	bridge	200	0.300	0.295	0.281	0.277	0.94		
			300	0.283	0.287	0.275	0.267	0.97		
9	300	contact	200	0.67	0.68	0.75	0.78	1.12		
			300	0.62	0.63	0.77	0.79	1.24	5.4	8.7
10	500	contact	300	0.48	0.44	0.69	0.72	1.44		
			500	0.50	0.46	0.74	0.69	1.48	6.8	13.6
11	500	bead	500	0.54	0.52	0.56	0.55	1.04	2.9	5.4
12	1,000	contact	600	0.223	0.207	0.222	0.228	1.00		
			1,000	0.206	0.185	0.196	0.204	0.95	0.93	4.5
13	5,000	contact	3,000	0.345	0.316	0.326	0.320	0.95		
			5,000	0.325	0.284	0.296	0.290	0.91		

ranged from 0.5 to 2 seconds. For converters of higher range, they were about 0.2 to 0.3 seconds.

Discussion

The results show that in general the time constants of converters may be expected to be dependent upon both the magnitude and direction of the step of heater current, with changes of 20% in some cases. Thus any specification of the time constant must include the conditions of test if accuracies better than this are desired

The measured time constants t_a and t_b generally differed from each other under the same test conditions by only a few per cent. Large differences, up to 50%, were obtained in a few cases, however. As a check these differences were verified for converters no. 6, no. 9 and no. 10 by carefully measuring the areas and ordinates on the oscillograms used for measuring t_a, and then calculating t_b. The calculated values agreed with the fluxmeter values of t_b to 3 per cent or better.

There are several causes of the depedence of both t_a and t_b upon the heater current, and also of the differences between t_a and t_b for the same current. They may be listed as follows:

1. Thomson effect in the thermocouple wires, which causes the emf versus temperature characteristic of the thermocouple to be nonlinear.
2. Nonlinearities caused by temperature dependence of the heating and cooling coefficients of the heater.
3. Temperature rise of the junctions between the lead-in-wires and the heater or thermocouple wires.
4. The fact that the heater is a complex thermal body with no single time constant.

These were investigated theoretically to determine the approximate magnitudes of their effects.

As shown in Appendix III, a nonlinear relation between the temperature rise and the output emf of the thermocouple of a converter can cause both t_a and t_b to be dependent on the magnitude and direction of the change in heater current. For the materials normally used in thermal converters, this effect would not be expected to cause t_a or t_b to differ by as much as 1%, and would cause either one to change by no more than 3 or 4%, as the step change in heater current is increased to rated value. In addition it would cause the change in time constant with current switched on to be equal and opposite to that with current switched off. This does not fit the observed patterns of Table I, for, in general, the low-current values are larger than those at rated current, with current switched either on or off.

Nonlinearities in the heater circuit can be caused by temperature dependence of either the heating or cooling coefficients. For most thermal converters heat is lost from the heater wire chiefly by conduction to the terminal blocks or lead-in wires. The change in temperature is thus a function of both time and the distance from the ends of the heater. Solutions for the resulting steady-state temperature distribution with nonlinear characteristics and with either direct or alternating current through the heater have been obtained.[3] Corresponding transient solutions should also be possible, but have not been attempted because of the complexity of the problem. Appendix IV gives an approximate solution for t_a and t_b for conductors with nonlinear lateral cooling only (convection and radiation). For these cases differences in t_a and t_b of almost 10% may be expected. Decreases in either t_a or t_b should be expected as the heater current is increased, with larger decreases (up to 70%) with the current switched on than with the current switched off. It is believed that similar patterns would be obtained for conduction cooling. Thus the nonlinearities in the thermocouple and heater account reasonably well in a semi-quantitative way for the observed typical dependence of the time constant of a converter upon the direction and magnitude of the change of heater current.

Neither of these effects can be expected to cause differences between t_a and t_b as large as those observed for converters no. 9 and no. 10 (20 to 50%). These differences are believed to be due to the temperature rise of the junctions between the heater and lead-in wires. Approximate calculations show that for these converters and for other glass-bulb converters of high-range (200–1,000 ma), the rise can be 5 to 10% of the temperature rise of the midpoint of the heater, with a time constant up to several times that of the heater wire alone. Formulas for t_a and t_b have been developed for the resulting case in which the change in emf is the sum of two unequal exponential terms, with results which seem in reasonable agreement with the observed differences.

Goodwin has shown[4] that the temperature rise of a conductor cooled by conduction to terminals at each end can be represented by a series of exponentials. The conductor is thus not a "simple thermal body." The series converges rapidly, however, and calculations indicate that the effect should cause t_a and t_b to differ by less than 1% and be independent of the current level.

Measurements showed that for all converters except no. 9 and no. 10 the ratio of the 90% response time, to the time constant, t_a, differed from that for a simple exponential (2.3) by less than 10%. For these it was 2.6 and 3.1 respectively. As shown in Table I, the ratio of the 99% time t_r to the time constant t_a differed appreciably from that for a simple exponential (4.6), for a large percentage of the few converters tested. Although additional converters should have been checked, this indicates that it may be desirable to directly measure rather than calculate t_r for those instrument applications in which this is the significant dynamic characteristic.

Summary and Conclusions

The change in emf of a thermal converter after a step change in heater current is not strictly exponential. Because of this two different definitions of an arbitrary time constant have been proposed for converters. One is the time required for 63% of the change in emf to take place, and the other is the time integral of the difference between instantaneous and final emfs divided by the total change in emf. Reasonably simple methods of measuring each have been demonstrated and the limitations on accuracy have been theoretically and experimentally determined. The results indicate that the two definitions generally (but not always) give the same result. For most of the 13 converters tested the differences were less than 5%, but for one the difference was almost 20% and for another 50%, under identical test conditions.

For most converters the measured time constants depend upon the magnitude and the direction (whether up or down) of the step change in heater current, with variations up to 10% rather common. Analysis indicates that these variations are very probably caused by nonlinearities in the heater and thermocouple. Discussions with manufacturers and users indicate that an accuracy of 10% is generally adequate in specifying time constants, although an accuracy of 5% is sometimes required. Because of the influences already noted, the time constant should be defined and the conditions of test should be specified. The author suggests a step change from zero to rated current as a standard condition in evaluating time constants of converters.

For many instrument applications the response time is of more importance than the time constant. For a thermal converter a definition of response time is suggested as the time required for 99% of the change in emf to take place after a step change from zero to rated current.

Because the same method and equipment may be used to measure the 63% and 99% response, the arbitrary 63% definition of "time constant" is believed by the author to be somewhat preferable to the definition based on the time integral of emf, if a choice must be made. (In a recent paper[5] the time for 63% change of response is also used as the criterion of the dynamic characteristics of thermocouples for temperature measurement, and is defined as the "characteristic time.") Calibration of the oscillograph would ordinarily not be necessary if an accuracy of 10% is sufficient, thus simplifying the test method considerably.

Inquiries indicate that the terms "time constant" and "response time" are often used loosely and even interchangeably, so that care must be taken in comparing or specifying values based upon them.

Appendix I. Error of Galvanometer Oscillograph with Exponential Forcing Function

The response of the idealized galvanometer (with a linear voltage amplifier of high input impedance) when the input voltage E decays exponentially with a time constant $t_c=1/k$, is given by the equation

$$P\frac{d^2\theta}{dt^2}+A\frac{d\theta}{dt}+U=E\epsilon^{-kt} \tag{1}$$

with the initial conditions $\theta_0=E/U$, and $d\theta/dt=0$ at $t=0$, where θ is the angular deflection, and P, A and U are the inertial, damping, and restoring constants respectively.[6]

For critically damped case, when $A^2=4PU$, the solution is

$$\theta=\frac{E}{Q}[\epsilon^{-kt}-(1-mt-kt)\epsilon^{mt}]+\frac{E}{U}(1-mt)\epsilon^{mt} \tag{2}$$

where $m=-A/2P=-2\pi/t_0=-\sqrt{U/P}$, $Q=Pk^2-Ak+U$, and t_0 is the free (undamped) period of the galvanometer. If the ratio n of the free period divided by the time constant is less than about 0.5, this becomes, closely enough, when $t=t_c$,

$$\theta_c=\frac{\theta_0}{\epsilon}\left(1+\frac{\eta}{\pi}\right) \tag{3}$$

Combining this with the approximate relation $\Delta\theta/\theta=-\Delta t/t_c$ shows that the error in measuring the time constant is about 30 t_0/t_c per cent. Since the period of the galvanometer used is about 0.02 second, an error of about $0.6/t_c$ per cent can be expected.

In general for $n<0.5$ and $t>t_c$ the response is $\theta=\theta_t\ (1+n/\pi)$ where θ_t is the response of galvanometer without error. Thus the error in measuring the 99% response time should be about $7t_0/t_c$ per cent, since $t_r=4.6\ t_c$.

It is tempting to consider measuring the response time of a converter by connecting it to a millivoltmeter and noting the time for 99% of the change in deflection to take place. While this gives a good measure of the minimum reading time of the combination, analysis shows that errors considerably larger than 10% would be expected for most converters ($t_c=0.2$ to 2 seconds), since the period of the usual sensitive millivoltmeter is about 2 seconds. This was borne out by a few informal tests of several converters with such a millivoltmeter.

Appendix II. Errors in Measuring t_b with a Fluxmeter

For an ideal fluxmeter having no restoring torque the equation of motion, neglecting inertial torque, is

$$A\frac{d\theta}{dt}=Gi \tag{4}$$

where i is the current through the moving coil and G is the motor constant. If an emf e is applied and the circuit is nonreactive, with a resistance R the resulting current is, $i=(e-e_b)/R$ where $e_b=Gd\theta/dt$. Thus

$$A'\frac{d\theta}{dt}=\frac{Ge}{R} \tag{5}$$

where

$$A'=A+G^2/R$$

When integrated from t_1 to t_2, this gives

$$A'(\theta_2-\theta_1)=\frac{G}{R}\int_{t_1}^{t_2}edt \tag{6}$$

without restriction on the duration of the time interval.

If $A\ll G^2/R$, the change in deflection is not dependent on R. In a shunted multirange fluxmeter, however, this is not usually the case.

There are several sources of error in using a fluxmeter in this way. If the errors are small, we may treat each source separately, since their combination will in general differ by second and higher order terms only.

The term $Pd^2\theta/dt^2$ must be added to equation 5 to account for the inertia of the moving system. However, if the moving system is stationary at the beginning and end of the time interval it is easy to show by direct integration that this causes no error. It can also be shown that no first-order error will be caused by the inductance of the circuit.

The limiting error for measurements at long time intervals is generally caused by the small restoring torque $U\theta$ of the suspensions (and in unbalanced instruments, by a gravitational couple). If the applied voltage e decays exponentially with a time constant $t_c=1/k$, we have

$$A'\frac{d\theta}{dt}+U\theta=\frac{GE\epsilon^{-kt}}{R} \tag{7}$$

with zero deflection defined as the rest position with no emf applied.

The solution of this equation, with $\theta=0$ at $t=0$, is $\theta=[S/(1-b)](\epsilon^{-bkt}-\epsilon^{-kt})$ where $b=U/kA'$ and $S=GE/kA'R$. When $b\ll1$ the maximum value of θ is approximately

$$\theta_m=S[1+b\ l_n\ b] \tag{8}$$

The value of b can easily be determined by observing the change in deflection from θ_0 to θ_1 in a measured time t_c, with $e=0$ and with the same circuit resistance. Then $A'd\theta/dt=-U\theta$, with $\theta=\theta_0$ at $t=0$, from which $b=\Delta\theta/\theta_0$ if $\Delta\theta=(\theta_0-\theta_1)\ll\theta_0$.

With an initial deflection of 60 divisions of the fluxmeter used in these tests, b was about 0.005 for $t_c=10$ seconds. Thus the restoring-torque error should be less than 3% for even the slowest converter tested.

Appendix III. Effect of Nonlinearity of the Thermocouple

When the heater current of a converter is switched off the temperature difference between the two junctions of the thermocouple may be represented by $\theta=\theta_0\epsilon^{-t/t_c}$. If the emf of the thermocouple is $e=\alpha\theta+\beta\theta^2$, with $\beta\ll\alpha$, direct substitution and integration gives

$$t_{b1}=\frac{1}{E}\int_0^\infty e\,dt=t_c\left(1-\frac{\beta\theta_0}{2\alpha}\right) \tag{9}$$

where E is the emf when $t=0$, and the subscript 1 refers to the condition, current switched off. When $e=E/\epsilon$ we find that, closely enough,

$$t_{a1}=t_c\left(1-\frac{0.6\beta\theta_0}{\alpha}\right) \tag{10}$$

When the current is switched on, we have $\theta=\theta_0(1-\epsilon^{-t/t_c})$ from which, in the same way

$$t_{b2}=t_c\left(1+\frac{\beta\theta_0}{2\alpha}\right) \tag{11}$$

and

$$t_{a2}=t_c\left(1+\frac{0.6\beta\theta_0}{\alpha}\right) \tag{12}$$

For the thermocouple materials ordinarily used in converters, $\beta/\alpha\approx+0.0004$ and $\theta_0\approx150$ C at rated heater current. Thus differences between t_a and t_b from this cause should not exceed 0.4% and both t_a or t_b should change by no more than 4% from low to rated current.

Appendix IV. Effect of Nonlinearities of the Heater

If we consider a conductor of uniform cross section, a, and circumference, c, heated by a steady current, I, suddenly applied, and with lateral cooling only, we have

$$asd\frac{d\theta}{dt}+h_0c\theta=\frac{I^2p}{a} \tag{13}$$

where d, s, and p are the density, specific heat, and resistivity of the conductor and h_0 is the cooling coefficient (watt/square centimeter degrees centigrade). If the initial conductor is $\theta=0$ at $t=0$ the solution is

$$\theta=\frac{g}{k_0}(1-\epsilon^{-k_0t}) \tag{14}$$

where $g=I^2p/a^2ds$, $k_0=h_0\ c/ads=1/t_c$, and t_c is the time constant of the temperature rise of the conductor.

The cooling coefficient h is in general temperature-dependent. For the moderate values of temperature rise in a converter we may approximate this by assuming that $h=h_0\ (1+\gamma\theta)$ where h_0 is the value at the reference temperature and γ is a constant. Temperature coefficients of the other quantities such as resistivity are in general less than γ for most converters. The differential equation then becomes

$$\frac{d\theta}{dt}+k_0\theta+\gamma k_0\theta^2=g \tag{15}$$

The effect of the small perturbing term, $\gamma k_0\theta^2$, which makes this a nonlinear equation, may easily be approximated by substituting in it the value of θ previously found. When this is done a linear equation is obtained. A solution for the two time constants is somewhat tedious but is straightforward. To a sufficiently close approximation for these purposes the results are

$$t_{a2}=t_c(1-1.4\gamma\theta_0) \quad t_{b2}=t_c(1-1.5\gamma\theta_0)$$

where the subscript 2 refers to the condition, current switched on, and

$$\theta_0=g/k_0$$

A similar procedure for the equation corresponding to current switched off gives

$$t_{a1}=t_c(1-0.6\gamma\theta_0)$$

and

$$t_{b1}=t_c(1-0.5\gamma\theta_0)$$

Values of about +0.003 might be expected for γ. Thus this effect would cause both time constants to decrease with increasing current, for both directions of the step change (current on and current off), with a greater decrease when current is turned on. Changes of up to 70% should be expected, considerably more than those observed.

In actual converters cooling is chiefly by conduction rather than convection and radiation, so that more complicated equations must be used. In addition some converters, particularly in the ampere ranges, have appreciable temperature coefficients of resistance. Nevertheless, the above approximations give what is believed to be a reasonable semi-quantitative picture of the situation.

References

1. Electrical Indicating Instruments. *ASA C39.1-1955*, American Standards Association, New York, N. Y., 1955.

2. The Inductronic Fluxmeter, R. W. Gilbert. *Weston Engineering Notes*, Newark, N. J., vol. 8, no. 1, 1953.

3. Thermal Converters as AC-DC Transfer Standards for the Measurement of Current and Voltage at Audio Frequencies, F. L. Hermach. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 48, 1952, p. 121 (RP 2296).

4. Heat Flow—An Important Factor in Electrical Design, W. N. Goodwin, Jr. *Technical Monograph TM-102*, Weston Electrical Instrument Corporation, Newark, N. J.

5. How to Specify Thermocouple Response, R. J. Moffat. *Journal*, Instrument Society of America, Pittsburgh, Pa., vol. 4, June 1957, p. 219.

6. Electrical Measurements (book), F. K. Harris. John Wiley & Sons, Inc., New York, N. Y., 1952, chap. 14.

◆

Discussion

W. C. Downing, Jr. (Sangamo Electric Company, Springfield, Ill.): Mr. Hermach refers to the use of the 90%-response time in specifying the dynamic characteristics of thermal converters as originating "perhaps because this is easier to measure." It is quite true that the 90% point can be determined much more easily and accurately than the 99% point but the source for this practice, I believe, is in the Code for Electricity Meters, where the time interval of a thermal demand meter is defined as the time required to reach 90% of ultimate reading on a steady load. While a thermal demand meter and a thermal converter have vastly different response times, the response characteristics are quite similar; and it has been natural, therefore, for manufacturers of both types of instruments to apply the same definition of response time.

For the determination of response time of thermal converters for the measurement of watts, we have found the use of a commercially available cathode-ray oscilloscope a straightforward technique that gives quite adequate results. With the oscilloscope set for linear time sweep on the horizontal axis, the converter output is connected, through a d-c preamplifier, to the vertical sweep circuit. By a double pole switch, load is applied to the converter and the time sweep triggered simultaneously. A single sweep is photographed, resulting in a plot, approximately 1 by 1.6 inches, of emf against time. For testing a watt-converter with 90% response in one second, a two second time sweep permits time determination of the 90% point with sufficient accuracy. Incidentally, since the watt-converter is designed for linear output and negligible temperature error, the response characteristic as determined in this manner closely approximates a true exponential curve.

John H. Miller and **John Nagy** (Weston Instruments, Newark, N. J.): The paper by Mr. Hermach in which he discloses the vagaries of the thermal converter with respect to the rising characteristic of its output when a step function of current is applied through the heater, is a valuable addition to the literature. These vagaries have been known for some years but, to the best of our knowledge, have never been delineated in detail. Attempts to state a single figure for the time constant of a thermocouple in simple terms have failed in the past and the reason for such failure is now obvious.

In view of the reported results it becomes quite apparent that no simple single figure will suffice for a statement of the time dependent output of one of these intriguing converting devices.

On the other hand, a simple method would be most desirable if it could be attained within the bounds of a reasonable specification.

Some recent work in the Weston Laboratories along the lines of matching the rise of the output millivoltage to the rise in an RC circuit, as shown in Fig. 4, has resulted in a modified circuit as shown in Fig. 5 of this discussion. Here we have a composite circuit having two superimposed RC networks, the first, R_1C_1, effectively delaying the current flow producing the opposing emf, and the second $(R_1+R_2)C_2$, producing an exponential rise.

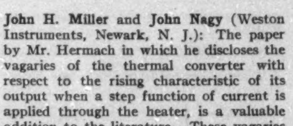


Fig. 5. Double mesh circuit for "bridge" method

To simplify the manipulation, R_3 and R_4, control the steady-state condition and where R_4 is kept low to have negligible effect on the timing circuit. Since the balancing drop in R_4 is usually less than 10 millivolts, it may be typically 1 ohm, so that for a dry cell R_3 is 150 to 1,000 ohms.

In the simple circuit of Fig. 4 of the paper, the balancing voltage of the RC network V_t at time t after the current is turned on, is given by

$$V_t=E\left(1-\epsilon^{\frac{-t}{R_1C}}\right) \tag{16}$$

and R_1C in ohms and farads is the time constant, t_a, for 63% of the change to take place.

In the composite circuit shown in this discussion a solution can be derived using the LaPlace transform method shown in Gardner and Barnes, "Transients in Linear Systems." The solution is

$$V_t=E\left(1+\frac{\Gamma\epsilon^{-\alpha t}-\alpha\epsilon^{-\Gamma t}}{\alpha-\Gamma}\right) \tag{17}$$

The exact values of α and Γ are somewhat complicated, but if R_2 is greater than 10 R_1 as is the case with those thermoelements we have tested in this manner, and if we make $C_1=C_2=C$, then, within 1% we have,

$$\alpha=\left(\frac{1}{R_1}+\frac{1}{R_2}\right)\frac{1}{C} \text{ and } \Gamma=\frac{1}{(R_1+R_2)C} \tag{18}$$

The characteristic time t_c derived in this manner, is, then,

$$t_a = \frac{1}{\alpha} + \frac{1}{\Gamma} = \frac{R_1R_2C}{R_1+R_2} + (R_1+R_2)C \qquad (19)$$

where we may call $1/\alpha$ the delay time, and $1/\Gamma$ the time constant of the exponential rise after the delay.

This can be further simplified, still within 1%, to

$$t_a = (R_2+2R_1)C \qquad (20)$$

In our study we used $C = 1{,}000$ microfarads, and for a typical 2.0-ma thermoelement, using an L&N Type *2430-C* galvanometer, we obtained $R_1 = 50$ ohms and $R_2 = 830$ ohms whereby $t_a = 0.93$ second. The balance obtained was excellent, the excursions each side of zero being less than 1/10 as much as with the simple circuit. The value given for t_a is for rated heater current turned "on." As shown by Hermach, different values were obtained for current turned "off," and still different values were obtained for other values of heater current.

In view of the findings reported in the paper and our own work, it is suggested that specifications for thermal converters must necessarily include the items listed in the following

For steady-state conditions, the following should be given:

1. Rated heater current, with tolerance, to produce rated millivolt output, measured on open circuit.
2. Resistance of heater, with tolerance, at current to produce rated millivolt output.
3. Resistance at output terminals of thermocouple with tolerance, when producing rated millivolt output.

When a specification for the output millivolt rise or delay time is a requirement, in addition to the foregoing, it is necessary to state the following items:

1. Environment, that is, ambient temperature and kind of enclosure, if any, of the thermoelement. Preferably the final form of mounting and enclosure should be used in making the test.
2. Are the data to be taken with current applied or removed?
3. State value of heater current.
4. State required percentage of steady state output millivolts increase or decrease, for heater current value used, for which elapsed time is required.

We feel that the term "characteristic time" used for the value of the time for 63% change in response has considerable merit and would urge that this term be used wherever appropriate. When values for the 90% or 99% change of response are required instead of this characteristic time, they should be so stated in the specification.

It is hoped that further consideration of these several factors may result in reasonable uniformity of specifications having to do with these converting elements to the end that their makers and users speak the same language.

F. L. Hermach: The information given by Mr. Downing regarding response time is appreciated. Since there is some precedent for it, and since it is much easier to measure, the 90% response time should probably ordinarily be specified rather than the time for 99% response.

Mr. Downing shows that a cathode-ray oscillograph is also satisfactory and convenient for measuring the 90% response time. The paper should have emphasized that any oscillographic method of suitable accuracy may be used to measure both t_a and response time, but not (directly at least) t_b.

The information given by Mr. Miller and Mr. Nagy is most interesting. We have also observed in our tests the small delay in the change of response of thermal converters. This delay requires that the true starting time be determined by some suitable marker on the oscillogram when an oscillographic method is used. The very small excursions of the ballistic galvanometer indicate that their double *RC* method can be used to measure the characteristic time t_a. As Miller and Nagy indicate, the complex formula for t_a becomes remarkably simple if $C_1 = C_2$ and $R_2 > 10R_1$.

The definite specifications given by Miller and Nagy should completely eliminate the present ambiguities pertaining to both the steady-state and dynamic characteristics of thermal converters. It is suggested that rated heater current suddenly applied (current on) be considered the standard test condition for 2 and 3 of their specifications for the transient test.

Minimization of Components in Electronic Switching Circuits

T. J. BEATSON
ASSOCIATE MEMBER AIEE

SWITCHING CIRCUIT SYNTHESIS presents two basic problems to the engineer. First, he must be able to obtain a circuit which will function according to the input-output specifications. Second, the circuit must be economical in the use of switching components, whether they are transistors, diodes, vacuum tubes, or relay contacts.

The first problem can always be solved. If the input-output specifications are in the form of a table of combinations, by treating one output at a time each output can be expressed as a Boolean function of the inputs. The Boolean function can then be translated into a circuit which will perform according to the given specifications. In most cases such a superficial treatment results in a circuit containing many unneeded circuit elements.

The distinction between different Boolean functions and different algebraic forms of the same Boolean function must be recognized. It is possible to manipulate a Boolean function into many different algebraic forms without changing the actual function. For every algebraic form of a function there is a corresponding circuit realization in the sense that the algebra describes the interconnection of circuit elements. Therefore the problem under consideration may be stated as follows: Given a Boolean function, find an algebraic form of this function, the circuit realization of which requires a minimum number of circuit components.

In this paper, minimization is considered only for those circuits in which diodes or transistors are used as switching elements. (The number of resistors required is not taken into account.) The treatment required for vacuum-tube circuits would be somewhat more complex because of the relatively large voltage drop between the plate and cathode and the resultant circuit limitations. The analysis presented here could be modified for use with relay contact circuits, but such a treatment would ignore the possibility of bridge circuits.

Circuit Realization

In electronic switching circuits nodal voltage is used as the characteristic property. The voltage at a given node is either high (1), or low (0), as determined by the current flowing into the node through a resistance connected to the supply voltage. Some typical diode and transistor switching circuits which are

---

Paper 58-3, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted July 26, 1957; made available for printing October 10, 1957.

T. J. BEATSON is with the General Electric Company, Palo Alto, Calif.

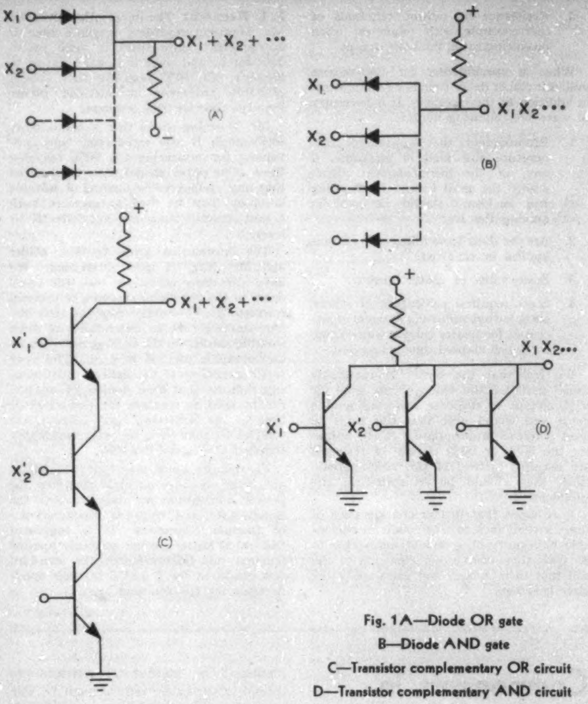


Fig. 1A—Diode OR gate
B—Diode AND gate
C—Transistor complementary OR circuit
D—Transistor complementary AND circuit

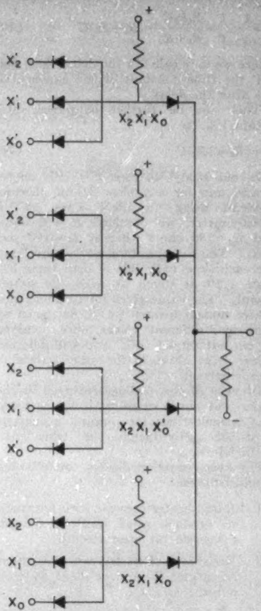


Fig. 2. Diode circuit corresponding to equation 1

operated by positive-going inputs are shown in Fig. 1. Either the diode gates of Fig. 1(A) or (B), or the transistor circuits of (C) or (D) may be cascaded, with the use of suitable voltage dividers, in order to construct the logic for a particular Boolean function.

Switching circuits, both combinational and sequential, are commonly specified by means of a table of combinations as shown in Table I. Each row of the table of combinations corresponds to a set of input states which must exist simultaneously in order that the output may have the indicated state. Every row of the table of combinations represents a different set of input states and the corresponding output state.

Table I. Table of Combinations

Inputs			Output
X_2	X_1	X_0	f
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	1

Let it be required to obtain a diode circuit having the logical structure of Table I. The Boolean function describing the logic of Table I is

$$f = X_2'X_1'X_0' + X_2'X_1X_0 + X_2X_1X_0' + X_2X_1X_0 \quad (1)$$

Without performing any algebraic manipulations of equation 1 the diode circuit corresponding to equation 1 may be constructed from the gates shown in Fig. 1(A) and (B). The resulting circuit is shown in Fig. 2.

If it is recognized that the last term equation 1 can be included in the function twice without altering the function, then equation 1 can be simplified by combining the second and fourth terms, and the third and fourth terms according to the theorems of Boolean algebra. Equation 2 is the result of this simplification.

$$f = X_2'X_1'X_0' + X_1X_0 + X_2X_1 \quad (2)$$

The diode circuit corresponding to equation 2 may be constructed in a manner similar to that used for Fig. 2. The result is shown in Fig. 3. This algebraic manipulation has resulted in a saving of six diodes, since both Figs. 2 and 3 represent different forms of the same function.

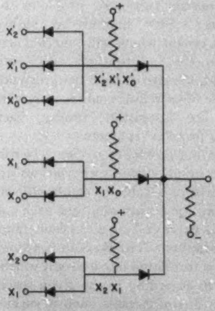


Fig. 3. Diode circuit corresponding to equation 2

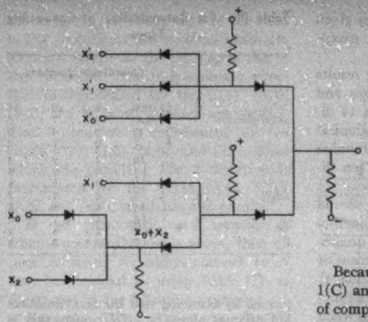


Fig. 4. Diode circuit corresponding to equation 3

Equation 2 may be further simplified by factoring X_1 from each of the last two terms, as shown in equation 3.

$$f = X_2'X_1'X_0' + X_1(X_0+X_2) \tag{3}$$

The circuit representing equation 3 is shown in Fig. 4, revealing an additional saving of one diode.

In each of the circuits of Figs. 2, 3, and 4, the number of diodes required is equal to the sum of the number of factors and the number of terms in the corresponding form of the Boolean function. In Fig. 4 and equation 3, for example, there are five factors and four terms as follows:

Factors	Terms
X_2'	$X_2'X_1'X_0'$
X_1'	$X_1(X_0+X_2)$
X_0'	X_0
X_1	X_2
(X_0+X_2)	

Although the foregoing example is by no means a proof of the fact, a given form of a Boolean function can always be realized as a diode circuit in which the number of diodes required is equal to the sum of the total number of factors and the total number of terms in the given form of the Boolean function. Table I may be realized as transistor circuits by following analogous reasoning with Fig. 1(C) and (D), with one additional step.

The circuit realization of the various terms and factors in a given form of a Boolean function may be systematically broken down into a number of stages. Thus in the first stage of equation 1 the individual terms are realized by taking various products of input variables. In the second stage of equation 1 the function is obtained by adding the various terms produced in the first stage. The outputs from the first stage are considered as inputs to the second stage.

Because the transistor circuits of Fig. 1(C) and (D) have the inherent property of complementing the input variables, it is necessary to take the complement of the variables at the inputs to every other stage in the algebraic form. In general, if the algebraic form consists of more than two stages (equation 3, for example) the following rule applies for the realization of transistor circuits:

> If the number of stages in an algebraic form of a Boolean function is even, then the inputs to all even-numbered stages must be complemented, along with the outputs from all odd-numbered stages. If the number of stages is odd, then the terms "odd" and "even" in the preceding statement should be interchanged.

Fig. 5 shows a transistor circuit obtained from the algebraic form of equation 3. Note the similarity between Figs. 4 and 5. Fig. 5 and equation 3 also illustrate the fact that a given form of a Boolean function can always be realized as a transistor circuit in which the number of transistors required is equal to the sum of the total number of factors and the total number of terms in the given form of the Boolean function. Therefore, for circuit realization using diodes or transistors the "minimal" algebraic form of any Boolean function is that form in which the total number of terms and factors is a minimum.

Boolean Algebra and Canonical Forms

Every Boolean function may be written in two different standard, or canonical, forms. These canonical forms serve as a convenient starting point for minimization since they may be readily obtained from a table of combinations.

The first canonical form of a Boolean function consists of a sum of terms and each term consists of a product of all variables, either primed or unprimed. Thus equation 1 is the first canonical

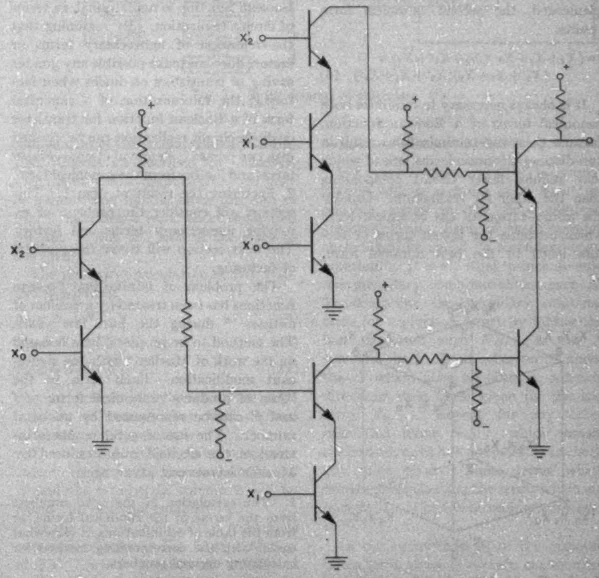


Fig. 5. Transistor circuit corresponding to equation 3

Table II. Table of Combinations

Inputs				Output
X_3	X_2	X_1	X_0	f
0	0	0	0	0
0	0	0	1	1
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	1	0

form of the Boolean function which represents the logic of Table I.

The second canonical form of a Boolean function consists of a product of factors, and each factor consists of a sum of all variables, either primed or unprimed. Again using the example of Table I, the second canonical form can be obtained by first writing the complement of the function (corresponding to the rows of Table I for which $f=0$) as in equation 4.

$$f'=X_2'X_1'X_0+X_2'X_1X_0'+X_2X_1'X_0'+X_2X_1'X_0 \quad (4)$$

When both sides of equation 4 are complemented, the second canonical form results.

$$f=(X_2+X_1+X_0')(X_2+X_1'+X_0)(X_2'+X_1+X_0)(X_2'+X_1+X_0') \quad (5)$$

It is always necessary to minimize both canonical forms of a Boolean function, because these two minimizations result in two distinct algebraic forms, one of which may require fewer circuit components than the other for realization. Usually the minimization of the canonical form which contains the fewest terms or factors leads to the best minimal form.

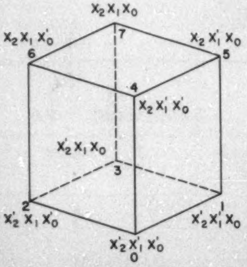


Fig. 6. Graphical representation of connections

However, the table of combinations given in Table II shows that there are exceptions to this rule.

The first canonical form which results from Table II contains seven terms and cannot be realized with less than 14 diodes or transistors. The second canonical form resulting from Table II contains nine factors and can be realized with ten diodes or transistors.

Any minimization applied to the second canonical form of a Boolean function (such as equation 5) can also be accomplished by minimizing the complement of the second canonical form (equation 4) since these two forms of the Boolean function are duals. This fact is useful when further minimization by factoring is attempted. For example, it is somewhat easier to factor equation 6

$$X_4X_3'X_2X_0'+X_4X_2X_1X_0 = X_4X_2(X_3'X_0'+X_1X_0) \quad (6)$$

than it is to factor the dual expression given by equation 7

$$(X_4'+X_3+X_2'+X_0)(X_4'+X_2'+X_1'+X_0') = X_4'+X_2'+(X_3+X_0)(X_1'+X_0') \quad (7)$$

Minimization of Canonical Forms Without Factoring

In general a canonical form of a Boolean function is nonminimal, in terms of circuit realization. By reasoning that the retention of unnecessary terms or factors does not make possible any greater saving of transistors or diodes when factoring, the minimization of a canonical form of a Boolean function for transistor or diode circuit realization can be divided into two parts: 1. Removal of unnecessary terms and factors from the canonical form. 2. Factoring the result of item 1. This section will consider the problem of removing unnecessary terms and factors. The next section will cover the problem of factoring.

The problem of minimizing Boolean functions has been treated by a number of authors[1–6] during the past few years. The method to be proposed here is based on the work of Mueller,[4] with one significant modification. Each term in the "sum of products" canonical forms of f and f' can be represented by an octal number. The use of octal numbers instead of the decimal numbers used by Mueller[4] has several advantages:

1. The calculation of the octal numbers from the terms of the canonical forms, or from the table of combinations, is somewhat easier than the corresponding method for calculating decimal numbers.

2. The necessary arithmetic calculations involving the octal numbers may be performed by assuming that the octal numbers are decimal numbers. (Of course, this is not true in general, but only for this particular numbering system and for the particular arithmetic calculations required by this method.)

3. A possible source of error in the decimal method arises because there is uncertainty as to whether to add or subtract a power of two from any given decimal number. This uncertainty is overcome in the octal method by referring to a short tabulation which contains all of the necessary information.

Table III. For Determination of Connecting Terms

		Co-ordinate Numbers		
Digit position	1,000	9	10	11
	100	6	7	8
	10	3	4	5
	1	0	1	2
Digit	0	1	2	4
	1	0	3	5
	2	3	0	6
	3	2	1	7
	4	5	6	0
	5	4	7	1
	6	7	4	2
	7	6	5	3

To calculate the octal number representation for a given term in the canonical form the variables in the term are first partitioned into sets of three. Each such set will determine one of the octal digits. All primed variables are given a weight of 0, while all unprimed variables are given a weight of 4, 2, or 1, depending on the position of the variable within the set.

Example: The octal number representing the term $X_7X_6X_5X_4X_3'X_2'X_1X_0'$ from a canonical form involving eight variables is obtained as follows:

Variables	X_7X_6	$X_5X_4X_3'$	$X_2'X_1X_0'$
Weights	2 1	4 2 0	0 2 0
Sums of weights	3	6	2

Thus the octal number 362 represents this term.

If two terms in a canonical form of a Boolean function are identical except for one variable (in general, the *i*th variable), it is possible to write the sum of these two terms as a single term. For example:

$$X_2X_1'X_0+X_2X_1X_0=X_2X_0 \quad (8)$$

Under these circumstances the two terms in question are said to be connected along the *i*th co-ordinate. Thus the two terms in equation 8 are connected along the 1st co-ordinate.

When the number of variables is five or less, it is possible to illustrate the connections between terms graphically. For three variables the terms can be considered as the vertexes of a cube, as shown in Fig. 6. Each edge of the cube in Fig. 6 connects two vertexes which are identical except for the priming of one of the variables. The 0 co-ordinate direction is represented by the (0,1), (2,3), (4,5), and (6,7), all of which are parallel. The 1 direction is represented by the edges (0,2), (1,3), (4,6), and (5,7), all of which are parallel. The 2 direction is represented by the edges (0,4), (1,5), (2,6), and (3,7), which are also parallel.

If one term with octal number N from a canonical form is given, then all possible terms N_i which connect to N may be obtained by using Table III, as illustrated in the following example:

Example: Given the term $N=362$ (octal) from a canonical form involving eight variables. Determine the eight terms which can be connected to 362. Those terms connected to 362 along co-ordinate directions 0, 1, and 2 are obtained by changing only the least significant digit of 362. This is indicated at the top of Table III. Since the least significant digit of 362 is 2, Table III shows that 2 should be changed to 3, 0, and 6. Thus in the 0, 1, and 2 co-ordinate direction 362 is connected to

$N_0=363/0$
$N_1=360/1$
$N_2=366/2$

Here the number to the left of the slant line is the octal number to which N is connected, and the number to the right of the slant line indicates the co-ordinate direction of the connection.

In the 3, 4, and 5 directions Table III indicates that the digit 6 in 362 should be changed to 7, 4, and 2. Thus 362 is also connected to

$N_3=372/3$
$N_4=342/4$
$N_5=322/5$

In the 6 and 7 directions Table III indicates that the digit 3 in 362 should be changed to 2 and 1. Thus 362 is also connected to

$N_6=262/6$
$N_7=162/7$

Mueller[4] has shown that an irredundant "sum of products" form of a Boolean function may be obtained from the first canonical form of the function by performing a series of simple arithmetic operations on the numbers which represent the terms of the canonical form. It should be clear that since the complement of the second canonical form of a given Boolean function is also the first canonical form of the complement of the given Boolean function this procedure may be applied to both the first canonical form and the complement of the second canonical form. The procedure described by Mueller, with the previously indicated change, may be summarized by the block diagram of Fig. 7, in which a work sheet with two columns at the left side of the sheet is assumed.

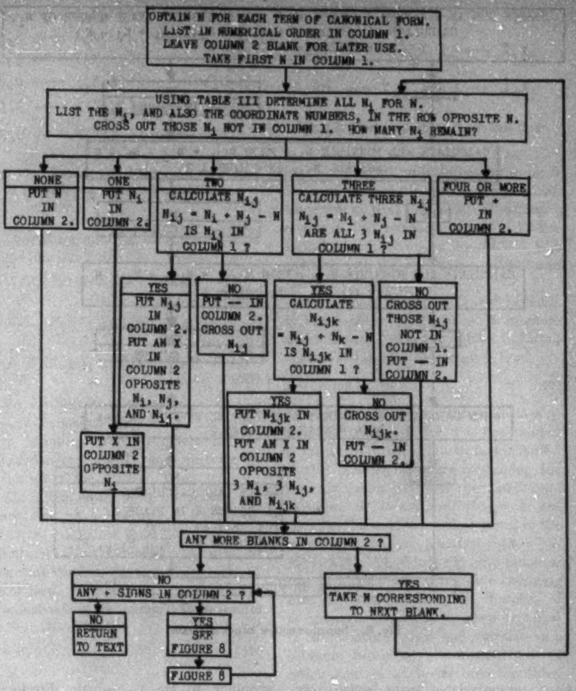

OBTAIN N FOR EACH TERM OF CANONICAL FORM.
LIST IN NUMERICAL ORDER IN COLUMN 1.
LEAVE COLUMN 2 BLANK FOR LATER USE.
TAKE FIRST N IN COLUMN 1.
USING TABLE III DETERMINE ALL N_i FOR N.
LIST THE N_i, AND ALSO THE COORDINATE NUMBERS, IN THE ROW OPPOSITE N.
CROSS OUT THOSE N_i NOT IN COLUMN 1. HOW MANY N_i REMAIN?
NONE
PUT N IN COLUMN 2.
ONE
PUT N_i IN COLUMN 2.
TWO
CALCULATE N_{ij}
$N_{ij} = N_i + N_j - N$
IS N_{ij} IN COLUMN 1 ?
THREE
CALCULATE THREE N_{ij}
$N_{ij} = N_i + N_j - N$
ARE ALL 3 N_{ij} IN COLUMN 1 ?
FOUR OR MORE
PUT + IN COLUMN 2.
YES
PUT N_{ij} IN COLUMN 2. PUT AN X IN COLUMN 2 OPPOSITE N_i, N_j, AND N_{ij}.
NO
PUT — IN COLUMN 2. CROSS OUT N_{ij}
YES
CALCULATE N_{ijk} $= N_{ij} + N_k - N$ IS N_{ijk} IN COLUMN 1 ?
NO
CROSS OUT THOSE N_{ij} NOT IN COLUMN 1. PUT — IN COLUMN 2.
PUT X IN COLUMN 2 OPPOSITE N_i
YES
PUT N_{ijk} IN COLUMN 2. PUT AN X IN COLUMN 2 OPPOSITE 3 N_i, 3 N_{ij}, AND N_{ijk}
NO
CROSS OUT N_{ijk}. PUT — IN COLUMN 2.
ANY MORE BLANKS IN COLUMN 2 ?
NO
ANY + SIGNS IN COLUMN 2 ?
YES
TAKE N CORRESPONDING TO NEXT BLANK.
NO
RETURN TO TEXT
YES
SEE FIGURE 8
FIGURE 8

Fig. 7. Block diagram of procedure

The numerical calculation of N_{ij}, N_{ijk}, and N_{ijkl}, from

$$N_{ij}=N_i+N_j-N \tag{9}$$

$$N_{ijk}=N_{ij}+N_k-N \tag{10}$$

$$N_{ijkl}=N_{ijk}+N_l-N \tag{11}$$

may be performed by treating the octal numbers as though they were decimal numbers; see Appendix I.

For those N which have a number entered in column 2 the combination of N and the number in column 2 will be called "an essential combination."

At the bottom of Fig. 7 it is sometimes necessary to refer to Fig. 8, an extension of Fig. 7 which requires no additional work sheet. Equations 9, 10, and 11 may be generalized so that

$$N_{ijklm\ldots rs}=N_{ijklm\ldots r}+N_s-N \tag{12}$$

For every N which has a (—) in column 2 there will be a choice of at least two possible terms in the final minimal algebraic form. Each of these possible choices will be called "a secondary combination." It is essential to include only one secondary combination for each N. To determine the secondary combinations for a given N, start by taking the last calculated term $N_{ij}\ldots$ in the N row and check off all other terms under this N which have no other co-ordinate directions than those given for the last term $N_{ij}\ldots$. Repeat for any other calculated terms $N_{pq}\ldots$ which remain unchecked, until it is not possible to check any more terms. Those terms which remain unchecked are the terms which are opposite to N in the secondary combinations.

In the event that all of the opposite terms for a given N contain one or more common co-ordinate directions, then an x should be placed in column 2 opposite

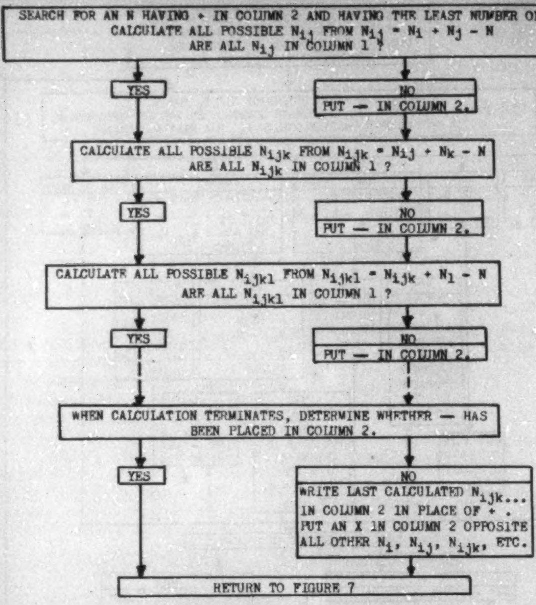

SEARCH FOR AN N HAVING + IN COLUMN 2 AND HAVING THE LEAST NUMBER OF N_i.
CALCULATE ALL POSSIBLE N_{ij} FROM $N_{ij} = N_i + N_j - N$
ARE ALL N_{ij} IN COLUMN 1 ?
YES
NO
PUT — IN COLUMN 2.
CALCULATE ALL POSSIBLE N_{ijk} FROM $N_{ijk} = N_{ij} + N_k - N$
ARE ALL N_{ijk} IN COLUMN 1 ?
YES
NO
PUT — IN COLUMN 2.
CALCULATE ALL POSSIBLE N_{ijkl} FROM $N_{ijkl} = N_{ijk} + N_l - N$
ARE ALL N_{ijkl} IN COLUMN 1 ?
YES
NO
PUT — IN COLUMN 2.
WHEN CALCULATION TERMINATES, DETERMINE WHETHER — HAS BEEN PLACED IN COLUMN 2.
YES
NO
WRITE LAST CALCULATED N_{ijk}... IN COLUMN 2 IN PLACE OF + . PUT AN X IN COLUMN 2 OPPOSITE ALL OTHER N_i, N_{ij}, N_{ijk}, ETC.
RETURN TO FIGURE 7

Fig. 8. Supplementary block diagram

all of the calculated terms having only these common co-ordinate directions.

The collection of all essential combinations and all secondary combinations corresponds to the prime implicants of Quine.[2] The selection of a minimal set of secondary combinations which, together with all of the essential combinations, corresponds to the selection of a minimal set of prime implicants, may be accomplished by using a procedure developed by McCluskey[6] and will be illustrated later. However, instead of forming a complete prime implicant table as described by McCluskey, it is necessary only to form a table consisting of secondary combinations and those N which are marked with a (−).

After all of the essential combinations and a minimal set of secondary combinations have been determined, the minimal algebraic form which results from the particular canonical form being considered may be constructed as follows:

For each combination, essential or secondary, write down the algebraic term corresponding to N from the canonical form. Eliminate from this term all of the variables which have subscripts corresponding to the co-ordinate direction numbers of the opposite term or term listed in column 2. The resulting term must appear in the minimal unfactored algebraic form of the original function. The following example illustrates the application of this procedure:

Example: Find the minimal unfactored forms of the function f, specified in Table II. The octal term numbers for the function f are determined by inspection of Table II. They are 1, 2, 4, 5, 6, 11, and 12.

Starting with the first number in this list, the numbers of those terms which connect to 1 are determined. Table III indicates that 1 is connected to 0, 3, and 5 in the 0, 1, and 2 directions respectively. The second digit of the number 1 (or 01) is 0, which Table III indicates should be changed to a 1, 2, or 4. Since this problem is concerned with only four variables, those octal numbers from 0 to 17 are of interest. Thus 1 is connected to 11 in the 3 direction. The connection may be summarized as follows: 0/0, 3/1, 5/2, 11/3. Now, it may be seen that 0 and 3 do not appear in column 1 of Fig. 9, so these must be crossed off the list of connections to 1. Now, with two connections remaining uncrossed equation 9 yields

$$N_{23} = N_2 + N_3 - N = 5 + 11 - 1 = 15 \qquad (13)$$

Listed with the other connections is 15/23; also note that 15 does not appear in column 1 of Fig. 9 and therefore 15 must be crossed off the list of connections to 1. Proceeding in the same manner for the remaining terms in column 1, Fig. 9 is obtained.

Thus there are two essential combinations of terms from the first canonical form, namely, (11,1) and (12,2). Three terms are marked with (−) and herefore a minimal set of secondary combinations must be selected. This is done following the procedure outlined by McCluskey.[6] It is noted first that there are four possible secondary combinations: (4,5), (4,6), (5,1), and (6,2). Each of these secondary combinations corresponds to the replacement of two terms in the canonical form by a single term with one less variable. A subset of these four secondary combinations must be chosen so that each of the terms 4, 5, and 6 from the canonical form is included in at least one of the chosen secondary combinations. To help accomplish this, a table of secondary combinations is formed as shown in Fig. 10. At the left of this

col. 1	col. 2					
1	x	~~0/0~~	~~3/1~~	5/2	11/3	~~15/23~~
2	x	~~3/0~~	~~0/1~~	6/2	12/3	~~16/23~~
4	-	5/0	6/1	~~0/2~~	~~14/3~~	~~7/01~~
5	-	4/0	~~7/1~~	1/2	~~15/3~~	~~0/02~~
6	-	~~7/0~~	4/1	2/2	~~16/3~~	~~0/12~~
11	1	~~10/0~~	~~13/1~~	~~15/2~~	1/3	
12	2	~~13/0~~	~~10/1~~	~~16/2~~	2/3	

Fig. 9. Determination of one unfactored minimal form of Table II

	4	5	6	
4,5	x	x		A
4,6	x		x	B
1,5		x		C
2,6			x	D

Fig. 10. Table of secondary combinations

table are listed the secondary combinations. The (—) terms are listed across the top of the table. The x's in the table indicate those columns which are included in each of the secondary combinations. A subset of rows from Fig. 10 must be chosen so that each column contains at least one x. For a small table of secondary combinations, such as Fig. 10, the calculus of propositions may be used to determine the minimal sets of secondary combinations, which will be denoted by g. As shown in Fig. 10 the rows will be denoted by A, B, C, and D. Thus the 4 column is covered by A or B, the 5 column is covered by A or C, and the 6 column is covered by B or D.

$$g=(A+B)(A+C)(B+D) \qquad (14)$$

When equation 14 is multiplied out the result is

$$g=AB+AD+BC \qquad (15)$$

Thus the three columns of Fig. 10 are covered by rows A and B, or rows A and D, or rows B and C. Since each of these coverings involves the same number of secondary combinations, any one of them can serve as the minimal set of secondary combinations. Reconstructing the terms which correspond to the essential and secondary combinations results in Fig. 11.

11,1 —— $\cancel{X_3}X_2'X_1'X_0$ —— $X_2'X_1'X_0$
12,2 —— $\cancel{X_3}X_2'X_1X_0'$ —— $X_2'X_1X_0'$
4,5 —— $X_3'X_2X_1'\cancel{X_0}$ —— $X_3'X_2X_1'$
4,6 —— $X_3'X_2\cancel{X_1}X_0'$ —— $X_3'X_2X_0'$
1,5 —— $X_3'\cancel{X_2}X_1'X_0$ —— $X_3'X_1'X_0$
2,6 —— $X_3'\cancel{X_2}X_1X_0'$ —— $X_3'X_1X_0'$

Fig. 11. Reconstructed terms corresponding to essential and secondary combinations

Thus three minimal unfactored forms of the function f are

$$f=X_2'X_1'X_0+X_2'X_1X_0'+X_3'X_2X_1'+X_3'X_2X_0' \qquad (16A)$$

$$f=X_2'X_1'X_0+X_2'X_1X_0'+X_3'X_2X_1'+X_3'X_1'X_0 \qquad (16B)$$

$$f=X_2'X_1'X_0+X_2'X_1X_0'+X_3'X_2X_0'+X_3'X_1X_0' \qquad (16C)$$

To consider the minimization of the second canonical form of f, the octal term numbers for f' are determined by inspection of Table II, and are listed in column 1 of Fig. 12.

Since the essential combinations are discovered before reaching 7, 10, 13, 16, or 17, no calculation is necessary for these numbers. Thus f' consists entirely of essential combinations and no table of secondary combinations is necessary. Reconstructing the minimal form results in Fig. 13. Thus

$$f'=X_2'X_1'X_0'+X_1X_0+X_3X_2 \qquad (17)$$

and

$$f=(X_2+X_1+X_0)(X_1'+X_0')(X_3'+X_2') \qquad (18)$$

Factoring

Although no systematic method is known of factoring Boolean functions to yield the minimal algebraic form, circuit components may be conserved by recognizing common factors when they appear in the minimal unfactored form of a Boolean function. For example, by recognizing that X_1 was a common factor in two terms of equation 2, the factored form given by equation 3 was written, with the resultant saving of one diode or transistor.

0,10 —— $\cancel{X_3'}X_2'X_1'X_0'$ —— $X_2'X_1'X_0'$
3,17 —— $\cancel{X_3'}\cancel{X_2'}X_1X_0$ —— X_1X_0
15,16 —— $X_3X_2\cancel{X_1}X_0$ —— X_3X_2

Fig. 13. Reconstruction of minimal form

As shown in Appendix II, the saving s in the number of diodes or transistors which is achieved by any given factoring operation may be determined from

$$s=(r-1)p+k-2 \qquad (19)$$

where r is the number of terms which contain the common factor, p is the number of variables in the common factor, and k is the number of terms containing the common factor with only one additional variable. It is assumed that r is not equal to the total number of terms in the function. For any practical case, $r\geq 2$, $p\geq 1$, $k\geq 0$. Therefore the only practical case in which s will be negative (meaning that factoring increases the number of components required) is when $(r,p,k)=(2,1,0)$. Hence when one variable is a common factor in two terms, both of which contain at least three variables, this common factor should not be factored.

Re-examining equation 2, it may be seen that X_1 is a common factor in two terms, each of which contains only one additional variable. Thus in this case, $(r,p,k)=(2,1,2)$. Substituting these values in equation 19 yields the result $s=1$.

If more than one factoring operation upon the minimal unfactored form of a Boolean function is possible, it becomes necessary to determine the order in which the factoring operations should be performed. No satisfactory process for the solution to this problem is known except for the cut-and-try process.

Consider the following function:

$$f=X_4'X_3X_2X_1'X_0'+X_4'X_3X_2X_1X_0+X_4X_3'X_2'X_1X_0+X_4X_3X_2'X_1'X_0'+X_4X_3X_2X_1'X_0+X_4X_3X_2X_1X_0' \qquad (20)$$

One possible way of factoring equation 20 is shown in equation 21.

$$f=X_3[X_2(X_4'X_1'X_0'+X_4'X_1X_0+X_4X_1'X_0+X_4X_1X_0')+X_4X_2'X_1'X_0']+X_4X_3'X_2'X_1X_0 \qquad (21)$$

To obtain equation 21 two successive factoring operations are performed.

col. 1	col. 2							
0	10	~~1/0~~	~~2/1~~	~~4/2~~	10/3			
3	17	~~2/0~~	~~1/1~~	7/2	13/3	17/23		
7	x							
10	x							
13	x							
14	x	15/0	16/1	10/2	~~4/3~~	17/01	~~11/02~~	~~12/12~~
15	16	14/0	17/1	~~11/2~~	~~5/3~~	16/01		
16	x							
17	x							

Fig. 12. Determination of second unfactored minimal form of Table II

First, $(r,p,k)=(5,1,0)$, from which $z=2$. Then $(r,p,k)=(4,1,0)$, from which $z=1$. Thus the total saving for this sequence of factoring operations is $2+1=3$.

Another possible way of factoring equation 20 is shown in equation 22.

$$f=X_4X_3(X_2'X_1'X_0'+X_2X_1'X_0+X_2X_1X_0')+X_4'X_3X_2(X_1'X_0'+X_1X_0)+X_4X_3'X_2'X_1X_0 \quad (22)$$

Here the two factoring operations were $(r,p,k)=(3,2,0)$, from which $z=2$, and $(r,p,k)=(2,3,0)$, from which $z=1$. Therefore the total savings in components is again $2+1=3$.

Still another way of factoring equation 20 is shown in equation 23.

$$f=X_3X_2(X_4'X_1'X_0'+X_4'X_1X_0+X_4X_1'X_0+X_4X_1X_0')+X_4X_3'X_2'X_1X_0+X_4X_3X_2'X_1'X_0' \quad (23)$$

Here the single factoring operation performed was $(r,p,k)=(4,2,0)$, from which $z=4$. Thus equation 23 appears to represent the minimal factored form of equation 20.

Summary

The circuit having a minimum number of diodes or transistors, and representing the logic of a given table of combinations, may be obtained as follows:

1. From the table of combinations obtain both the first canonical form and the complement of the second canonical form of the Boolean function.
2. Eliminate redundant terms and factors from each of these forms by following the procedure outlined in the section "Minimization of Canonical Forms Without Factoring."
3. Remove common factors from each of the forms as indicated in the section "Factoring."
4. Count the total number of terms and factors in each of the two resulting forms and select that form of the Boolean function for which this count is a minimum.
5. Synthesize a circuit having the logic of the form of the Boolean function selected as the minimal form. (If this minimal form has been obtained from the complement of the second canonical form then the minimal form must be complemented before synthesizing the circuit.)

Appendix I. Manipulation of the Digits of Octal Numbers as Decimal Digits

Let A be a number. Then the digits of the octal representation of A are given by

$$A=A_{n-1}8^{n-1}+A_{n-2}8^{n-2}+\ldots+A_18^1+A_0 \quad (24)$$

Let equation 24 now be written as

Table IV. Conditions for Addition and Subtraction of Powers of 2

n	1	2	4
0	+1	+2	+4
1	−1	+2	+4
2	+1	−2	+4
3	−1	−2	+4
4	+1	+2	−4
5	−1	+2	−4
6	+1	−2	−4
7	−1	−2	−4

$$A=A_{n-1}10^{n-1}-A_{n-1}(10^{n-1}-8^{n-1})+A_{n-2}10^{n-2}-A_{n-2}(10^{n-2}-8^{n-2}+\ldots) \quad (25)$$

Rearranging equation 25

$$A=A_{n-1}10^{n-1}+A_{n-2}10^{n-2}+\ldots+A_110^1+A_0-A_{n-1}(10^{n-1}-8^{n-1})-A_{n-2}(10^{n-2}-8^{n-2})-\ldots-A_1(10^1-8^1) \quad (26)$$

It may now be seen that the quantity

$$\sum_{i=0}^{n-1}A_i(10^i-8^i) \quad (27)$$

represents the error which results when the digits of the octal representation of A are treated as though they were the decimal digits of another number A'.

Now let x, a, b, and c be numbers which satisfy the equation

$$x=a+b-c \quad (28)$$

It will be assumed that the octal representations of x, a, b, and c each consist of n digits. If the digits of the octal representations of a, b, and c are treated as though these digits were decimal digits then the resulting error on the right side of equation 28 will be

$$\sum_{i=0}^{n-1}(a_i+b_i-c_i)(10^i-8^i) \quad (29)$$

where a_i, b_i, and c_i are the octal digits of a, b, and c.

If instead the octal representations of a, b, and c are treated correctly as octal numbers, then the octal digits of x will be given by

$$\begin{aligned}
x_0&=a_0+b_0-c_0-8p_0\\
x_1&=a_1+b_1-c_1+p_0-8p_1\\
x_2&=a_2+b_2-c_2+p_1-8p_2 \qquad (30)\\
&\vdots\\
x_{n-1}&=a_{n-1}+b_{n-1}-c_{n-1}+p_{n-2}-8p_{n-1}\\
x_n&=p_{n-1}
\end{aligned}$$

In equation 30 p_i denotes the carry from the ith digit to the $(i+1)$th digit. Now, if the resulting octal digits of x are treated as though they are decimal digits then the resulting error in x will be

$$\sum_{i=0}^{n}x_i(10^i-8^i) \quad (31)$$

The x_i in equation 31 are the octal digits of x given by equation 30.

If the expressions in equations 29 and 31 are equal, no actual error results from treating the octal digits of a, b, and c as decimal digits instead of octal digits. Inspection of equations 29, 30 and 31 shows that if $p_i=0$ for all i, then equations 29 and 31 will be equal.

It now remains to be shown that for the octal representation of the numbers N_i, N_j, and N in equation 9

$$N_{ij}=N_i+N_j-N \quad (9)$$

there is no carry between digits ($p_i=0$). This means that

$$0\leq(a_i+b_i-c_i)\leq7 \quad \text{for } 0\leq i\leq n-1 \quad (32)$$

From the manner in which N_i and N_j are obtained from N by means of Table III it is evident that

$$N_i=N\pm2^i \quad (33)$$

In equation 33 the plus sign will apply if the expansion of N into powers of 2 does not contain 2^i. If the expansion of N does contain 2^i then the minus sign is used in equation 33. Similar considerations apply to the plus or minus signs used with N_j. Thus equation 9 may be written as

$$N_{ij}=N\pm2^i\pm2^j \quad (34)$$

In the octal system the powers of 2 are 1, 2, 4, 10, 20, 40, 100, 200, 400, etc. For any particular digit, n, of the octal representation of N this means that the only numbers which will be added to or subtracted from n are 0, 1, 2, or 4. Table IV shows the conditions for addition and subtraction of powers of 2 for any digit value n. This table also shows that for any possible digit n it will be impossible to add or subtract powers of 2 in such a way as to cause a carry to occur. Therefore no error results from treating the digits of the octal numbers in equation 9 as though they were decimal digits. By extending equation 28 it is possible to prove similar results applying to equations 10 and 11.

Appendix II. Saving of Components by Factoring

Assume that an unfactored Boolean function consisting of a sum of products is given. Let

s = number of terms in the Boolean function
m_i = number of variables in the ith term
q = number of terms in which $m_i=1$
r = number of terms containing a particular common factor
p = number of variables in the common factor
k = number of terms containing the common factor and for which $m_i=p+1$
x = number of terms plus the number of factors in the unfactored form of Boolean function
y = number of terms plus the number of factors in the factored form of the Boolean function
$z=x-y$ = saving of components due to factoring

Because the individual terms of the unfactored function may be arranged in any order, they can be arranged so that the first r terms contain the common factor. Then

$$x=(s-q)+\sum_{i=1}^{s}m_i \quad (35)$$

$$y=(s-q-r)+\sum_{i=r+1}^{s} m_i+(p+1)+r+\sum_{i=1}^{r}(m_i-p)-k+1 \quad (36)$$

$$z=x-y=(r-1)p+k-2 \quad (37)$$

References

1. Synthesis of Electronic Computing and Control Circuits, **Harvard University Staff.** "Annals of the Computation Lab. No. 27," Computation Laboratory, Harvard University Press, Cambridge, Mass., 1951.

2. The Problem of Simplifying Truth-Functions, **W. V. Quine.** *The American Mathematical Monthly*, Menasha, Wis., vol. 59, Oct. 1952, pp. 521–31.

3. The Map Method for Synthesis of Combinational Logic Circuits, **M. Karnaugh.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 593–99.

4. On the Synthesis of a Minimal Representation of a Logic Function, **Rolf Mueller.** Communications Laboratory, Air Force Cambridge Research Center, Cambridge, Mass., Apr. 1955.

5. A Topological Method for the Determination of the Minimal Forms of a Boolean Function, **R. H. Urbano, R. K. Mueller.** Computer Laboratory, *Ibid.*, Mar. 1956.

6. Minimization of Boolean Functions, **E. J. McCluskey, Jr.** *Bell System Technical Journal*, New York, N. Y., Nov. 1956, p. 1417.

◆

Discussion

Oscar Lowenschuss (Sperry Gyroscope Company, Sperry Rand Corporation, Great Neck, N. Y.): The paper contains two significant contributions: the minimization of canonical forms with the use of octal number representation, and a discussion of factoring and its resulting circuit economies.

The minimization of canonical forms with the use of octal numbers appears to offer several advantages over Mueller's decimal method, and over the ternary number method of Harris,[1] which was not mentioned in the paper. The discussion of factoring leads to an important relation, indicating possible component savings resulting from various methods of factoring. Unfortunately, Mr. Beatson's method does not, in general, lead to the most economical transistor circuit; in fact, his examples do not represent minimal circuits. Thus, if his equation 23 is written as

$$f=X_3X_2[X_4+(X_1X_0+X_1'X_0')][X_4'+(X_1X_0+X_1'X_0')']+X_2'X_4(X_3+X_1X_0)(X_3'+X_1'X_0') \quad (38)$$

the resulting circuit requires only 27 transistors instead of the 32 he mentions.

The factoring methods of Mr. Beatson use only the Boolean connectives "union" (inclusive or) and "intersection"; the "prime" (complement) has not been used. However, the prime, together with the "implicit function" concept, can lead to a more economical circuit realization, as shown by equation 38. It might be argued that the prime cannot be realized with diodes; however, it may be realized with a single transistor.

Implicit functions may be defined as follows: If a Boolean function $G(y_1, y_2, \ldots y_n)$ can be substituted into a given function $f(X_1, X_2, \ldots X_m, y_1, y_2, \ldots y_n)$ so that the function $F[X_1, X_2, \ldots X_m, G(y_1, y_2, \ldots y_n)]$ is equivalent to the given function f, then G represents an implicit function of f. Implicit functions may be detected by examining truth tables using a method due to Ashenhurst[2] or by "partial differentiation" by the method of Akers.[3] Thus in equation 33 the implicit function is

$$G(X_0,X_2)=X_0+X_2 \text{ and}$$
$$f(X_0,X_1,X_2)=f[X_1,G(X_0,X_2)]=X_1G(X_0,X_2)+X_1'G'(X_0,X_2)$$

It is noted that implicit functions provide circuit economies in a single-output switching circuit only if the implicit function appears in both primed and unprimed form. (In multiple-output switching circuits, the implicit function can provide a "kernel" that may help to minimize the entire circuit, even if it appears only in unprimed form.) Hence, one attempts to factor the given function f in one of the two forms

$$f=aG+bG'+c$$

or

$$f=(d+G)(e+G')h$$

where the letters a, b, c, d, e, and h represent variables, or functions of variables, provided only that any variables in G do not appear in a, b, c, d, e, or h. If G is a function of two or more variables, then a saving of circuit components is possible.

Minimization of switching circuits may now be divided into three parts:

1. Removal of unnecessary terms and factors from the canonical form, by the method given in this paper or equivalent methods.
2. Searching for an implicit function that appears in both primed and unprimed form, and contains at least two variables.
3. Factoring the result of item 2.

The author implies the following important result that when transistors are used to synthesize a switching circuit, the transistor configuration is essentially identical to the diode-gate configuration (except for the priming of certain input variables), as shown by a comparison of Figs. 4 and 5. It should be pointed out that the following restrictions must be observed:

1. If the complementary OR circuit, Fig. 1(C), is used exclusively, the transistor configuration is similar to the diode configuration only if the diode configuration ends in an OR gate, i.e., if the function has been factored in the first canonical form (sum of products). If the function has been factored in the second canonical form (product of sums), then the transistor circuit output must be complemented, or the states 0 and 1 must be inverted.

2. If the complementary AND circuit, Fig. 1(D), is used exclusively, the diode and transistor configurations are similar only if the diode configuration ends in an AND gate, i.e., if the function has been factored in the second canonical form. If the function has been factored in the first canonical form, the transistor circuit output must be complemented, or the states 0 and 1 must be inverted.

3. The similarity between diode and transistor circuits holds only if every diode AND gate drives only OR gates, and vice versa, as is usually the case. If a diode AND (OR) gate drives another AND (OR) gate, then in the transistor circuit an inverter must be interposed between such gates.

References

1. An Algorithm for Determining the Minimal Representations of a Logic Function, **B. Harris.** *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-6, June 1957, pp. 103–08.

2. The Decomposition of Switching Functions, **R. L. Ashenhurst.** International Symposium on the Theory of Switching, Harvard University, Cambridge, Mass., Apr. 2, 1957.

3. On a Theory of Boolean Functions, **S. B. Akers, Jr.** *Bulletin*, American Mathematical Society, Providence, R. I., vol. 63, Nov. 1957, p. 350.

E. J. McCluskey, Jr. (Princeton University, Princeton, N. J., and Bell Telephone Laboratories, Inc., Whippany, N. J.): Two interesting ideas are presented in this paper: first, the use of octal numbers for minimization of Boolean functions and, second, the formula for the saving in the number of diodes or transistors which is achieved by any given factoring operation. In connection with the minimization of Boolean functions, I should like to point out that use of a table such as that presented in my paper[1] eliminates the necessity of computing a list of the numbers of all terms which connect to the term being considered. Also, the use of Table III is unnecessary since the only computations required are additions of 1, 2, or 4 to a single digit of the number representing the term being studied.

In his discussion of minimal factored forms of Boolean functions, the author does not take into account the fact that it may be impossible to derive the minimal factored form directly from the minimal sum of products form. This is discussed in an article by William Burkhart,[2] in which one of the examples given is the function

$$F=(A+D)(B+CE)+DE'$$

References

1. See reference 6 of the paper, p. 1423.

2. Theorem Minimization, **William Burkhart.** *Proceedings*, Association for Computing Machinery, New York, N. Y., May 2–3, 1952, p. 259.

T. J. Beatson: I should like to thank the discussers for their contributions. Mr. Lowenschuss' discussion shows that the detection of implicit functions must be included as part of any comprehensive minimization technique, and is a very useful addition to this paper.

My discussion of factoring states specifically that no method of obtaining minimal factored forms is known, other than the trial-and-error method. Usually there is some amount of insight associated with a trial-and-error procedure, so that if one tries long enough and hard enough he should be able to arrive at the particular factored form suggested by Mr. McCluskey.

Measurement of Voltage Resulting From Single-Phase Switching of a High-Voltage 3-Phase Transformer

W. F DUNKLE
FELLOW AIEE

W. F. MACKENZIE
MEMBER AIEE

SINGLE-PHASE SWITCHING of small high-voltage transformers with large turns ratios can result in excessive secondary voltages and damage to connected equipment unless these conditions are recognized and measures taken to mitigate these phenomena. Failure of switchgear connected to the low side of a 66,000/480 500-kva delta-delta transformer prompted investigation and measurement of voltages attained on two similar installations. Measurements were found to be in reasonable agreement with those determined analytically.

The problem is divided into two parts:

1. The first problem involves the steady-state 60-cycle voltage that appears across the secondary terminals. This voltage results from the secondary winding capacitance to ground being energized in series with the interwinding capacitance by the 39-kv-to-ground voltage of the system when one or two primary fuses are removed. The solution of this problem requires only a low-impedance circuit connecting the secondary delta to ground. Methods used for analyzing the problem were substantiated by the investigations.

2. The second problem involves high-frequency transient oscillations which occur when a primary fuse is being either removed or replaced, and pronounced arcing occurs to the fuse disconnect. It is believed that this phenomenon may also cause secondary flashovers.

Calculation of Capacitances

When the secondary of a small transformer is ungrounded, the secondary voltage under certain conditions may be calculated by using the relative impedances of the various winding capacitance and the primary voltage.

Data from periodic insulation power factor tests made in accordance with AIEE Publication No. 505[1] facilitated determining the winding capacitances of the transformer. This calculation was premised upon the winding reactances to ground being primarily capacitive.

Insulation Field Test Data

No. of test	Circuit	Test Kv	Ma*
1	High side energized, low side grounded	10	11.15
2	Low side energized, high side grounded	2	4.08
3	High and low side tied together	2	3.6

* Milliamperes. Referring milliamperes obtained on tests 2 and 3 to 10 kv, and writing the capacitive currents in terms of the component values obtains the following simultaneous equations:

$$I_H+I_{HL}=11.15 \quad (1)$$

$$I_L+I_{HL}=20.4 \quad (2)$$

$$I_H+I_L=18.0 \quad (3)$$

Subtracting equation 1 from equation 2, adding this result to equation 3, and solving the capacitive components when calculated, give the following values:

$$I_L=13.7 \text{ ma} \quad (4)$$

$$I_H=4.3 \text{ ma} \quad (5)$$

$$I_{HL}=6.7 \text{ ma} \quad (6)$$

Since the assumption was that all the impedance was capacitive reactance

$$E=I_z \times X_c$$

or

$$X_c=\frac{E}{I_z}$$

and using the values from equations 4, 5, and 6 for determining capacitance impedance values.

$$\text{Total } X_{c3}=X_{CL}=\frac{E}{I_L}=\frac{10,000}{0.0137}$$
$$=730,000 \text{ or } 0.73 \text{ megohm}$$

$$\text{Total } X_{CL}=X_{CH}=\frac{E}{I_H}=\frac{10,000}{0.0043}$$
$$=2,320,000 \text{ or } 2.3 \text{ megohms}$$

$$\text{Total } X_{C2}=X_{CHL}=\frac{E}{I_{HL}}=\frac{10,000}{0.0067}$$
$$=1,490,000 \text{ or } 1.49 \text{ megohms}$$

Calculation of Voltage

The calculations of phase-to-ground secondary voltages on the station service transformer at Hummelstown Substation were made, with the use of the high-side winding capacitance to ground, the interwinding capacitance, and the low-side winding capacitance to ground. They are shown schematically in Fig. 1.

To make these calculations, the assumptions were made that: 1. the winding capacitances were lumped together, 2 the capacitances were connected to the midpoint of the transformer windings, and 3 the winding capacitances were symmetrical and equal in value for each phase. Fig. 1 shows these assumptions diagrammatically. Voltage values calculated were for steady-state conditions only, and they may be considerably higher during transient conditions existing while the transformer primary is being switched.

The steady-state primary phase-to-ground voltages and also the primary mid-winding-to-ground voltages are shown in Fig. 2(A). The shape and size of the secondary delta is fixed by the winding ratio and cannot be changed. The secondary delta is centered around ground by assuming C_2 equal for each phase, and C_3 equal for each phase. Under normal operation the C_3-to-ground voltage would be 139 volts, while the C_2 voltage is 19.8 kv.

One Phase Switched Out

When one primary fuse is removed, Fig. 3, single-phase operation results, and the de-energized terminal assumes a voltage approximately midway between the energized phases. Since the voltages are fixed by the winding ratios, full voltage will be across one winding while half voltage will be across the other two windings. The voltages to ground of the secondary will depend on the relative magnitudes of C_2 and C_3. With 19.9 kv

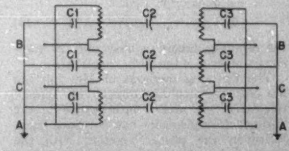


Fig. 1. Interwinding and winding to ground capacitance of 69/0.48-kv 500-kva transformer

---

Paper 58-35, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 10, 1957; made available for printing November 4, 1957.

W. F. DUNKLE and W. F. MACKENZIE are with the Pennsylvania Power & Light Company, at Hazleton, Pa., and Allentown, Pa., respectively.

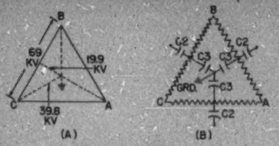


Fig. 2. Schematic representation of steady-state condition

A—Steady-state voltages to ground
B—Windings and capacitances

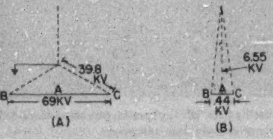


Fig. 3. Diagram showing voltages when one high-side fuse is removed

A—Primary
B—Secondary

across the capacitors C_2 and C_3, the secondary phase to ground voltage will be

$$V_{ABC} \text{ to ground} = \frac{0.73 \text{ megohms}}{2.22 \text{ megohms}} \times 19.9 \text{ kv} = 6.55 \text{ kv}$$

where 0.73 megohm is X_{C3} and 2.22 megohms is $X_{C3}+X_{C2}$

Two Phases Switched Out

In the case of Fig. 4, with two primary fuses removed, all terminals of the delta are at the same potential as the energized phase. This results in full 39-kv phase-to-ground voltage being impressed across C_2 and C_3 rather than the mid-winding voltage to ground as in the previous case. The secondary phase-to-ground voltage will therefore be

$$V_{ABC} \text{ to ground} = \frac{0.73}{2.22} \times 39.8 \text{ kv} = 13.1 \text{ kv}$$

Field Tests: Series 1

On December 14, 1954, tests were conducted on the 69,000/480-volt 500-kva station service transformer at Hummelstown Substation. The purpose of these tests was to simulate high-voltage conditions which had been experienced on switchgear connected to the 480-volt winding of this transformer and to obtain data for comparing calculated secondary phase-to-ground voltages against actual voltages obtained.

The tests were conducted under adverse weather conditions with rain falling continuously throughout the day. All measurements were taken at a switch box in the 66-kv yard with the secondary disconnect switch open. A 20-megohm resistance voltage divider was used to obtain voltage readings with a Triplett multimeter connected to a 2-megohm tap and ground. Test conditions were set up by removing and replacing primary fuses from the transformer with a switch stick as used under normal operating conditions.

On the first series of tests, voltages of 380, 100, and 340 volts, phase-to-ground, were obtained on the open secondary phases with C-phase primary fuse removed. These voltages are for A, B and C phases respectively. With B- and C-phase primary fuses removed, voltages of 600, 360, and 360 volts, phase-to-ground, were obtained on the open secondary phases. Fig. 5 showing connection of instruments. Preliminary calculations showed that these voltages should have been about 6 kv and 13 kv for the two setups respectively. Visual investigation as to the cause of the low-measured voltages as compared to the calculated voltages led to the conclusion that the asbestos base-fiber panel of the switch mount had absorbed moisture under the damp conditions prevailing, and that sufficient was leaking through the panel to reduce the voltages to the values obtained.

All three primary fuses were then removed, and a solid ground was connected to C-phase secondary terminal on the transformer. When A-phase primary fuse was inserted in its holder, and B and C primary phases open, readings of 370, 0, and 0 volts, phase-to-ground, were obtained on the open secondary phases. When an attempt was made to put B-phase primary fuse into its holder with A-phase primary fuse already inserted, unusually long arcs were noticed from the free end of the fuse to the fuse holder, and a flashover was observed to have occurred in the switch box. Upon removal of the switch panel the B-phase contact mounting screw was found to have flashed over to the metal switch box which was at ground potential. The clearance at the point of flashover between the contact mounting screw and the switch box was approximately 3/8 inch with sealing wax insulation between the two metals. The flashover appeared along the fiber mounting panel between the sealing wax and the mounting panel.

To secure better insulations to ground, a 1/4-inch Micarta panel was attached to the back of the switch-mounting panel, and the switch was remounted in the switch box. The first series of tests with the ground removed from the secondary were repeated. With one phase energized, the voltage measured on all phases to ground was approximately 10,000 volts. With two phases energized, voltage measurements on all three phases to ground were approximately 5,000 volts. During these tests, corona formations were noticed on the switch contacts and also between the panel-mounting screws and the panel board. Because of the adverse weather conditions, tests were discontinued. It was believed that adequate information had been obtained to substantiate the preliminary calculations.

Field Tests: Series 2

On January 12, 1955, a second series of tests were conducted on a similar 69,000/480-volt 500-kva station service transformer at Hosensack Substation. The objective was to compare data on this transformer with results obtained on the transformer at Hummelstown Substation.

The test setup and measuring circuit were identical to those used at Hummelstown. A cathode-ray oscilloscope was connected to one of the secondary terminals through a resistance voltage divider. Both transformers were similar as to voltages and ratings but of different manufacture, and the transformer secondary switches were of different design.

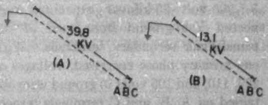


Fig. 4. Schematic representation showing voltages to ground with two high-side fuses removed

A—Primary
B—Secondary

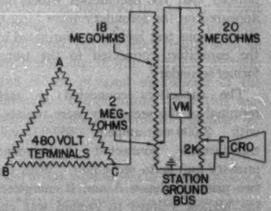


Fig. 5. Schematic of connections for measuring voltages to ground from transformer 480-volt terminals

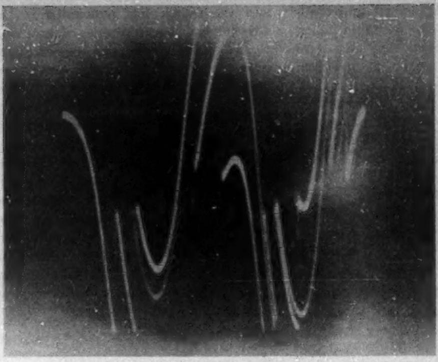


Fig. 6. Cathode-ray oscillogram of C-phase-to-ground voltage observed with A- and C-phase primary fuses removed

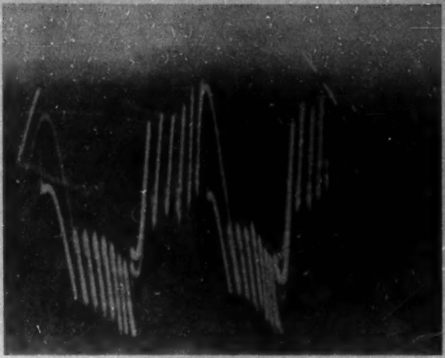


Fig. 7. Cathode-ray oscillogram of C-phase-to-ground voltage with C-phase primary fuse removed

With B primary phase energized and the secondary open, voltages of 750 volts to ground were measured on each phase. A considerable number of oscillations on the secondary wave form in the steady-state condition were noticed with this setup. A photograph of C-phase secondary voltage to ground was taken with the cathode-ray oscillograph, which is shown in Fig. 6. When A- and B-phase primaries were energized, voltages of less than 750 volts were obtained on each phase to ground. Fig. 7 shows this secondary voltage to ground on C phase.

These tests were then repeated with 3-7,200-volt 25-kilovar capacitors connected Y-to-ground from each of the transformer secondary terminals. With one primary phase energized, voltages of 220, 110, and 105 volts to ground were obtained on A, B, and C terminals respectively. With A- and B-phase primary phases energized, voltages of 145, 320, and 205 volts to ground were obtained. With all three primary phases energized, voltages of 260 volts to ground were obtained on all three secondary phases. It was noticed that the oscillations which had appeared on the secondary wave form in the previous test had been eliminated by the capacitors connected to the transformer secondary.

The capacitors connected to each terminal of the secondary were replaced with 3-500-volt-ampere 460/115-volt transformers connected Y-to-ground primary and delta secondary. With the two primary phases A and B energized, the highest voltage to ground was obtained on B phase, which was about 70 volts. With only B-phase primary energized, voltages of 87, 223, and 138 volts were obtained on A, B, and C secondary phases respectively. With all three primary phases energized, voltages of 260 volts to ground were obtained on all three phases. It was also noted that with this transformer type of grounding, the oscillations on the secondary wave shape were eliminated. With A-phase fuse in position and while the B-phase fuse was being put in its holder, B phase flashed over in the secondary switch mechanism.

Examination of the secondary switch disclosed that a Micarta insulating tube of approximately 1/32-inch thickness which fitted over a horizontal fastening bolt on B pole of the switch was missing, which reduced ground clearance at the point of flashover to about 0.125 inch.

Analysis of Field Test Results

As far as the 60-cycle voltages on the secondary are concerned, it appears that the connection of either transformer or capacitor to the secondary can hold the steady-state voltages to reasonable value. However, analysis of the condition under which flashovers occurred during switching indicate that the transformer is not adequate to reduce voltages during transients, but that capacitors are effective.

As a result of the tests, 2,000-volt oil-impregnated 10-microfarad capacitors were connected on both Hummelstown and Hosensack transformer secondaries. At each location capacitors were connected in Y between each phase of the transformer terminals and ground. The addition of these capacitors has entirely eliminated the high voltage resulting from single-phase switching.

From the tests the following preliminary conclusions were obtained:

1. An asbestos-base, fiber panel which is used to mount the secondary disconnect switch absorbs moisture, thereby reducing its electrical insulating properties.
2. Abnormally high voltages exist on the secondary terminals when single-phase switching is done under normal operating procedure.
3. The 60-cycle voltages which appear on the secondary during switching may be calculated from the interwinding capacitances.
4. The connection of the capacitors to the secondary minimizes the 60-cycle or transient voltage on the secondary during switching.
5. The connection of transformer to the secondary appears to minimize only the 60-cycle voltages on the secondary winding during switching.

Reference

1. PROPOSED TEST CODE FOR POWER FACTOR TESTING OF POWER TRANSFORMERS. *AIEE Publication no. 505*, Jan. 1955.

◆

Discussion

S. Bennon (Westinghouse Electric Corporation, Sharon, Pa.): The authors have made a timely contribution by calling attention to the possibility of low-frequency overvoltages due to capacitance effects. The possibility of overvoltages for impulse waves has received considerable attention in the past, which might give the impression that the problem is associated only with lightning surges. However, as demonstrated by the authors, for certain types of excitation and for certain classes of transformers overvoltages can be produced at low frequency.

As the authors also indicate, voltage unbalance is necessary. Balanced 3-phase excitation applied to the primary of a 3-phase transformer with balanced winding capacitances will not displace the secondary. However, unbalanced excitation such as that produced by opening one or two lines will tend to cause winding displacement with respect to ground when the secondary is isolated.

These effects are most likely to be observed on transformers with low-voltage windings well below the 15-kv class. Above the 15-kv class the ratio of test voltage to operating voltage is set by the standards at a value of 2, but below the 15-kv class this ratio increases considerably. As a result, very low voltage windings, especially if they are overinsulated, will tend to have a lower capacitance to ground in proportion to their operating voltage than higher voltage windings. Thus, the percentage of the unbalanced voltage applied to the high voltage appearing between the low-voltage winding and ground will tend to increase. While these voltages will, in most cases, be within the insulation strength of the windings involved, they still introduce problems such as shock hazard and possible damage to associated equipment.

P. L. Bellaschi (Consulting Engineer, Portland, Oreg.): The paper exemplifies the importance that capacitance coupling between low- and high-voltage windings in transformers can have on the voltages developed in the low voltage. Under certain conditions, discussed in the paper, the voltage developed in the low-voltage winding can assume dangerously high levels sufficient to damage the insulation.

The high voltages developed in the low-voltage winding basically result from the high-capacitance coupling between the high- and low-voltage windings. Note that the ratio of the test levels for the two windings is not proportionately as high as the corresponding ratio of the capacitance coupling.

For instance, in the case cited, the electrical clearances from the 66,000-volt winding to the 480-volt would be in the order of one inch and from the 480-volt to the core or ground about 1/4 inch or less. Therefore, the induced voltage in the 480-volt winding is in the order of one fourth or thereabouts the voltage to ground of the high-voltage winding.

These overvoltages may be accentuated further as the result of switching operation. As pointed out in the paper, "capacitance loading" or various other loadings of the secondary will reduce the overvoltage to acceptable levels.

Difficulties of similar nature have been experienced in 3-phase transformers from impulse voltages induced into low-voltage tertiaries from high-voltage Y-connected windings. In some instances the cause of the failure could be ascribed to inadequate insulation or low insulation level and to floating tertiary. The paper amply illustrates how small things may develop consequential troubles.

J. B. Hodtum (Allis-Chalmers Manufacturing Company, Pittsburgh, Pa.): This extremely interesting paper brings up several ideas that may well prove to contribute to the voltage situation. For example, in a small unit the high-voltage bushing capacitance to ground may be such that when only one phase is excited, two phases of the winding can go into ferroresonance with the bushing capacitance. The question of ferroresonance, however, was not mentioned.

Recently, a considerable number of overvoltage troubles have been experienced on ungrounded delta systems, many of these on the 480-volt delta ones. Such systems have no fixed relation to ground except the inherent capacitance of the circuit itself. Should arcing ground occur on one phase, the other two phases result in an inductance-capacitance circuit due to the line capacitance to ground and the inductance of the circuit itself, and it may result in overvoltages. The installation of 25-kilovar capacitors on such a system may react as stabilizing means as long as arcing ground cannot occur on one phase of the circuit. Under arcing ground conditions the capacitors may exaggerate the voltage difficulties.

The use of potential transformers connected Y-delta with the Y point grounded acts as a limited grounding transformer. It therefore has the ability to fix the voltage to ground and prevent the accumulation of energy in the capacitance part of the circuit. However, should a fault occur on one of the legs of such a circuit, the potential transformer would be subjected to short circuit and would fail. It would be our recommendation that Y-delta grounding banks be used and that resistors be inserted in the neutral to limit the current to the rating of the grounding transformer bank. This current need only be enough to ensure that enough resistance is inserted in the inductance-capacitance circuit to prevent oscillation when arcing grounds occur.

W. F. Dunkle and W. F. Mackenzie: We are in full agreement with Mr. Bennon's statements and appreciate his substantiation of our thinking. Mr. Bellaschi's comments are also appreciated since they also confirm and substantiate our conclusions. We thank Mr. Hodtum for his comments and agree that the additional aspects of the problem which he mentions merit further investigation. However, utility funds and equipment to engage in this type of investigation are somewhat limited and this investigation was terminated when a practical solution to the problem was found. The use of Y-delta grounding transformers as a solution to this problem was found by test to be unsatisfactory, since this type of connection did not prevent a secondary flashover when arcing occurred on switching the second 66-kv phase. The reason for this secondary flashover was not explored.

Some Practical Mathematical Techniques

GORDON K. CARTER
MEMBER AIEE

THE various papers which have considered the subect of automation of engineering computations have shown that there are certain mathematical procedures, such as interpolation in tables and solution of equations, which occur with considerable frequency. In this paper an effort is made to give some more explicit information on a few of these procedures. It is not feasible to do more than indicate some of the methods which have been of value, but it is hoped that these may help to develop interest in a most fascinating and useful field. The way to further study is indicated by the references.

Although the discussion which follows is nominally separated into several discrete sections, there is considerable overlap of these topics both in theory and in practice. The reader is urged to approach them with the intention of understanding their principles, rather than of using them to classify any given computation in handbook fashion. Generally speaking, the successful application of numerical techniques is pretty much of an art, requiring understanding and patience; and the best method to use sometimes differs for apparently small differences between cases or problems. Mathematically, this may lead to long and elegant dissertations. But, from the practical standpoint, success often results from an understanding of the physical problem and of the physical significance of the associated computations. For this reason, it would be useful to visualize the

---

Paper 57-194, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 9, 1956; made available for printing December 6, 1957.

GORDON K. CARTER is with the General Electric Company, Schenectady, N. Y.

Table I. Data from Load Test

Current, Amperes	Terminal Voltage
0	220
10	217.4
20	214.3
30	210.3
40	204.7
50	183.2

examples in terms of physical situations with which one is personally familiar.

Curve Fitting

Exact Fitting of Tabulated Data

Tables of data are especially familiar in the form of function tables, such as sine or logarithm, and in the form of measured quantities, such as speed, current, or power. Among the most common operations on tabulated data are interpolation, integration, and differentiation. All such operations require something which the table itself is incapable of furnishing, namely, (1) the assurance that the table does in fact represent a (sufficiently) continuous function between the table entries, and (2) knowledge of how the function does in fact vary between the listed points. The table itself cannot give any information except what is in the table.

What reopens the door, thus unceremoniously slammed, is, of course, a knowledge of the function. It is known that the sine is a "smooth" function of the angle, or speed is a "smooth" function of the load. Of course, what is meant by a "smooth" function cannot be precisely explained but the meaning is clear. For instance, based on the data of Table I from a load test of a d-c shunt generator, the curve of Fig. 1 would appear to be a pretty good representation of this function. This is probably because a knowledge of the problem indicates that no humps are to be expected, and hence the points look close enough together to give good definition of the curve as drawn. Since the table did not show all this, though, it is clear that any automatic manipulations of the table must be made with due regard for the relationship of those manipulations to the function which they are supposed to approximate.

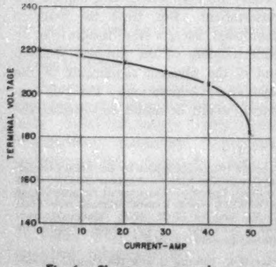


Fig. 1. Shunt-generator voltage

Linear interpolation is familiar to everyone. This is the equivalent of writing a linear equation, such as

$$V=a+bI \tag{1}$$

or

$$I=A+BV \tag{2}$$

to approximate the function. The coefficients are determined by causing the line to pass through the table points adjacent to the interpolated point; but, unless the equation of the line is particularly needed, this is not usually carried out explicitly, since it is shorter to interpolate by proportional parts. The linear function is the same whether it is as used in the form of equation 1 or of equation 2; and its relationship to the graph of the tabulated points is obvious, being the line segment joining the two adjacent points of the graph as in Fig. 2. These particular points have been chosen purposely, to show a rather poor approximation for the linear function.

If it is not possible, or convenient, to determine table values at a closer spacing, a closer approximation can be attempted by using higher degree polynomials, for instance, quadratic fitting:

$$V=a+bI+cI^2 \tag{3}$$

or

$$I=A+BV+CV^2 \tag{4}$$

The coefficients are determined by causing the quadratic to pass through the three points nearest the interpolated point. But now the two forms given in equations 3 and 4 are not usually equivalent, since equation 3 represents a parabola with vertical axis, and equation 4 one with horizontal axis. The only way to tell which is better is to see which agrees better with the smooth curve. This is generally true for nonlinear fitting. In the region shown in Fig. 2, for example, quadratic interpolation by equation 3 could be carried out by substituting the three corresponding sets of data into equation 3 and solving

$$\begin{cases} a+30b+900c=210.3 \\ a+40b+1{,}600c=204.7 \\ a+50b+2{,}500c=183.2 \end{cases}$$

This gives

$a=131.7$, $b=5.005$, $c=-0.0795$

and the result is shown in Fig. 3. The approximation is apparently better than linear over the right half of the region, but worse over the left half. The principal reason for this poor showing is the fact that the smooth curve is vertical, or nearly so, at 50 amperes; and polynomials of the type of equation 3 have difficulty in approximating this condition without producing bulges somewhere else.

Equation 4 could be fitted in the same way as equation 3, by solving suitable equations. Instead, it will be used as an example of the Lagrange method of writing the expression directly:

$$I=30\times\frac{V-204.7}{210.3-204.7}\times\frac{V-183.2}{210.3-183.2}+40\times\frac{V-210.3}{204.7-210.3}\times\frac{V-183.2}{204.7-183.2}+50\times\frac{V-210.3}{183.2-210.3}\times\frac{V-204.7}{183.2-204.7}$$

The form of this expression should be carefully noted, since it enables one to write a polynomial of any desired degree, through a corresponding set of points, with no more difficulty than shown here. Each term on the right consists of a value of the dependent variable (I), and a set of multiplying factors which take on the value 1 for the corresponding value of the independent variable (V) and the value 0 for each other value of the independent variable for the points used. This curve has not been plotted, but a few check points indicate that it gives a better fit than the straight lines in both halves of the range of Fig. 2.

This example has illustrated two ways of fitting a polynomial curve to given data points. The fitted curve, if it appears satisfactory, can then be used to estimate the value of the function between the table points (interpolation). Alternatively, the fitted curve can be integrated to give an estimate of the area under the function, or it can be differentiated to give an estimate of the derivative of the function. The validity of any of these operations cannot be assumed without suitable

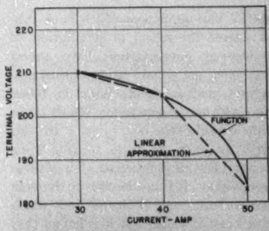


Fig. 2. Linear approximation

Vol.
Pt. 1
1958

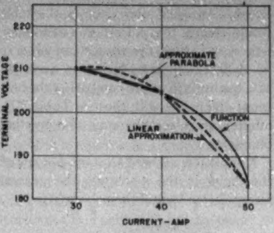


Fig. 3. Quadratic approximation

tests of the goodness of the fitted curve; however, it is generally the case that the integral tends to be better, and the derivative worse, than the quality evidenced by the fit of the function itself.

Because of the difficulty of checking fit, the exact fitting process is actually seldom applied to measured data. It is, however, extensively used for interpolation in function tables. It is also used in developing formulas for approximate differentiation and integration, especially for use in the numerical evaluation of integrals which cannot be formally integrated, and for use in the numerical solution of differential and integral equations which cannot be solved directly. Forms of other than polynomial type are normally avoided, unless there is a strong physical reason for introducing them.

APPROXIMATE FITTING OF TABULATED DATA

Especially for measured data, it is usually both appropriate and advisable to approximate the unknown function without attempting to fit the table points exactly. In this way it is frequently possible to obtain a suitably smooth approximating function which, at the same time, embodies the principal trends indicated by the tables. It is not too unusual in this case to make use of function forms other than the polynominal, such as exponential or sinusoidal forms, for example. This depends again upon knowledge of the physical situation, and it will not be pursued further here.

Whatever the form selected, the simplest case is that in which the dependent variable is assumed to be an explicit function of the indepedent variable. And the simplest treatment is to assume the values of the independent variable are true, while the values of the dependent variable are to be approximated. For example, consider the measured magnetic pull data of Table II, assuming the current values are correct. An attempt will be made to approximate the force using for the function

$$F=aI^2+bI^4 \qquad (5)$$

on the basis that the force should be zero for zero current, and that it should be an even function of current (same value for negative as for positive currents). Two conditions must be met. First, there must be more entries in the table than the number of unknown coefficients; this can be taken care of by using all four entries of the table to determine the two unknown coefficients. Second, there must be a reasonable criterion for good fit, one that makes sense and also provides a workable means of computing the unknown coefficients. Of those which are most often used, the best for many purposes is the least-squares criterion, used here for illustration.

Table II. Magnetic Pull Data

Current, Amperes	Force, Pounds	Force (Curve)
1	1	0.891
2	3	3.353
3	7	6.757
4	10	10.05

In view of the fact that equation 5 is not expected to be capable of fitting all the table points exactly, the error which it gives at each point of the table is noted:

$$\begin{cases} \epsilon_1=a+b-1 \\ \epsilon_2=4a+16b-3 \\ \epsilon_3=9a+81b-7 \\ \epsilon_4=16a+256b-10 \end{cases}$$

Under the least-squares criterion, the coefficients a,b are to be chosen such that the sum of the squares of all these errors is as small as possible

$$\Sigma\epsilon_i^2=(a+b-1)^2+(4a+16b-3)^2+ \\ \ldots+(16a+256b-10)^2=\text{minimum}$$

This is a polynomial (in fact, quadratic) function of a and b, and its minimum point will in general be such that the partial derivatives of the function with respect to a and b are zero; hence, with a factor of 2 removed,

$$\begin{cases} (a+b-1)+4(4a+16b-3) \\ \qquad +\ldots+16(16a+256b-10)=0 \\ (a+b-1)+16(4a+16b-3) \\ \qquad +\ldots+256(16a+256b-10)=0 \end{cases}$$

These equations suffice to determine a and b:

$$a=0.9081 \qquad b=-0.01748$$

The resulting curve has the values listed in Table II and the shape shown in Fig. 4, and would seem to be a sufficiently good representation for most manipulative purposes.

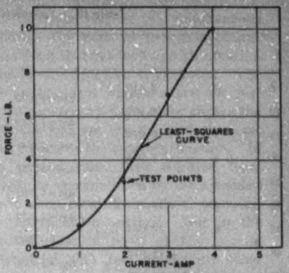


Fig. 4. Electromagnet pull test

The foregoing procedure, which is really quite simple in principle, is capable of application without much difficulty to almost any function form, so long as it is only the coefficients which need to be determined; thus, considerable latitude is available in the choice of function. If the number of table points just equals the number of coefficients, the result will be an exact fitting; but the labor will be greater than if the fitting is carried out in direct fashion.

The least-squares criterion is not necessarily the best for approximate fitting, although it is probably the simplest to use. The criterion that the sum of the absolute values of the errors be a minimum is sometimes preferred. Another useful one is that the magnitude of the largest error be as small as possible. In addition, it is possible to give the table values different weights, so that those which are considered most reliable have the most influence.

APPROXIMATE FITTING OF KNOWN FUNCTIONS

Surprising as it may seem, there is fairly frequent occasion to use an approximate representation of a function which is of known form. A familiar example would be the approximation of a rectified sine wave of voltage by means of a limited number of terms of its Fourier series. A less familiar, but very useful, case is the approximate representation of a segment of a sine function, say, by a finite polynomial expression so as to economize computing time with an automatic computer. Such approximations of known functions may permit the fitting criterion to be applied continuously over the whole interval, by integration, instead of by summation at discrete points as in the preceding section. The principles involved are just the same.

Table III. Calculation of Finite Differences

x	n	y				
5	0	2,119				
			$\Delta y_0=2,020$			
10	1	4,139		$\Delta^2 y_0=-89$		
			$\Delta y_1=1,931$		$\Delta^3 y_0=6$	
15	2	6,070		$\Delta^2 y_1=-83$		$\Delta^4 y_0=2$
			$\Delta y_2=1,848$		$\Delta^3 y_1=8$	
20	3	7,918		$\Delta^2 y_2=-75$		$\Delta^4 y_1=-3$
			$\Delta y_3=1,773$		$\Delta^3 y_2=5$	
25	4	9,691		$\Delta^2 y_3=-70$		$\Delta^4 y_2=1$
			$\Delta y_4=1,703$		$\Delta^3 y_3=6$	
30	5	11,394		$\Delta^2 y_4=-64$		$\Delta^4 y_3=-1$
			$\Delta y_5=1,639$		$\Delta^3 y_4=5$	
35	6	13,033		$\Delta^2 y_5=-59$		$\Delta^4 y_4=-2$
			$\Delta y_6=1,580$		$\Delta^3 y_5=3$	
40	7	14,613		$\Delta^2 y_6=-56$		
			$\Delta y_7=1,524$			
45	8	16,137				

It is interesting to find that the Fourier series approximation as usually described is, in fact, a least-squares fitting. The fact that the coefficients of the terms can be computed one by one, instead of by solving simultaneous equations as was done with the polynomial, is a result of the fact that the Fourier series components are orthogonal over the usual range of representation. This is an extremely valuable property for computation purposes since it permits direct observation of the improvement provided by each added term of the approximation.

A number of other families of orthogonal functions have found use in practical problems; e.g., the Chebycheff polynomials. The successive members of this family are polynomials of degree 0, 1, 2, and so on. They have the property that when a given function is approximated by a limited number of such terms, over the proper range, the maximum error of the approximation is the least that is possible by the use of a polynomial of that degree. For many purposes, as already mentioned, this kind of fitting is more advantageous than the more usual least-squares type.

Finite Differences

Another item to be considered under the general subject of curve fitting is the calculus of finite differences. In its most common (but not most general) form, this is a study of the properties of a polynomial which is fitted exactly to a set of data tabulated at equal intervals of the independent variable. These properties are explored in terms of the increments of the dependent variable, also called the "differences" of the function, whence the title. There is a very extensive literature on this topic. The general nature of it is a development in close analogy with that of differential calculus, with the differences corresponding to derivatives, summations to integrals, a general interpolation formula to Taylor's series, and difference equations to differential equations. Not only are the two systems formally similar in many ways, but also formal operational equivalences are developed which make it possible, for example, to write an expression for any specified derivative in terms of the differences, or conversely. As already mentioned, these matters are important in the development of approximate computing methods. They also turn out to be a useful means of studying the properties of certain types of systems of simultaneous equations arising from periodic structures or periodic switching phenomena.

While the subject cannot be studied extensively here, its nature can be indicated by exhibiting one form of the general interpolation formula, corresponding somewhat to Taylor's series in the calculus. The information in Table III starts with the independent variable x and dependent variable y, the former being taken at equally spaced locations.

The column headed "n" is a serial numbering of the x values to provide an equivalent variable with unit spacing; the location of the zero will be referred to later. The increments of y are written next to the spaces between y values and are designated here by one of several standard notations, Δ being read as "first forward difference of" the quantity following it. The subscripts correspond to the n's and would be changed if the location of the zero of n were shifted. The first differences constitute a tabular function also, and have increments or differences of their own. These are the "second forward differences of," denoted by Δ^2. And so on as far as desired. The fact that the actual numbers taper off and become substantially constant after a few differences is an indication that the tabulated function is closely representable by a low-degree polynomial in the range listed. An accidental error in the table would change this pattern in a characteristic way (which the reader can observe by deliberately introducing an error of, say, ten units in y_4 and comparing the new set of differences with those in Table III), offering an interesting way of correcting blunders.

Now, if n is allowed to stand for any of the digits in the n-column, the general interpolation formula may be written

$$y_n=y_0+n\Delta y_0+\frac{n(n-1)}{2!}\Delta^2 y_0+\frac{n(n-1)(n-2)}{3!}\Delta^3 y_0+\ldots \quad (6)$$

It is easy to verify, at least for the first few values of n, that this formula gives the numbers in the table. While not exactly the same as Taylor's series, it is clearly closely analogous in form, as already suggested.

A close look at Table III shows that Δy_0 depends only on y_0 and y_1; $\Delta^2 y_0$ only on y_0, y_1, and y_2; and so on. Thus if a definite number of terms is taken on the right side of equation 6, the formula involves exactly that number of terms from the table; and the formula will be found to give back only those same terms as values of n are inserted in it. So far, so good; nothing is obtained from the formula except the information that was put into it.

The next step is to observe that, if one now imagines n in equation 6 as a continuous variable, this expression becomes a polynomial in n of degree one less than the number of terms selected, and hence of degree one less than the number of points which it passes through. It is therefore *the* polynomial of lowest degree passing through the chosen points; and one has an alternative curve fitting scheme which is in many ways more convenient than either the substitution or the Lagrange method, although limited to evenly spaced points.

To be specific, from Table III,

$$y_n=2,119+2,020n$$

for the linear function through the first two points, or

$$y_n=2,119+2,020n-\frac{89}{2}n(n-1)$$

for the quadratic through the first three points, and so on. These functions can be used in all respects as before, for interpolation, for differentiation, for integration, or other desired purposes. The fact that they are expressed in terms of n has at least two features which need observation. First, best accuracy is provided by so placing the zero of n that the region of interest (interpolation point, derivative location, or region of integration) is centrally located with respect to the table points entering into the formula. Thus, to interpolate near $x=25$, using a second-degree formula, the zero of n is placed opposite $x=20$. Second, once the zero of n is located, it is a simple matter to determine the equivalence between x values and n values; this is required for describing the point at which interpolation or some other activity is to take place. The scale factor for differentiation or integration depends only on the x-spacing, since, for instance, from equation 6

$$\frac{dy_n}{dx}=\frac{dy_n}{dn}\frac{dn}{dx}=\frac{1}{\Delta x}\frac{dy_n}{dn} \qquad (7)$$

where Δx has the obvious interpretation.

In practice, the formulas developed by such methods as indicated are often translated back from the difference notation and given directly in terms of table entries. This frequently gives the best form for automatic computation; and it costs very little in the way of effort, since there are very simple relationships expressing differences in terms of table entries.

Solution of Nonlinear Equations

In engineering problems, solution of nonlinear equations or systems of nonlinear equations is a very common necessity. Frequently the need arises from trying to use interpolation formulas, or approximate fitted curves, in reverse. For instance, one might need to find the load current corresponding to a given terminal voltage from equation 3, or the current required to produce a given force in equation 5. A more complex case would arise, for instance, in trying to solve the heat balance relations for a turbine and its auxiliaries.

In the case of a simple interpolation function of one variable, as in the examples given here, one possibility is to set up the inverse interpolation function by simply interchanging the role of the variables. This would not, however, generally give the same results as the forward formula, as already noted. In simple cases the inverse of the forward interpolation function can be obtained though, and this is a desirable solution. Both equations 3 and 5, for instance, can be solved by the quadratic formula. Other simple solvable forms will occur to the reader, such as a single exponential or a simple trigonometric form.

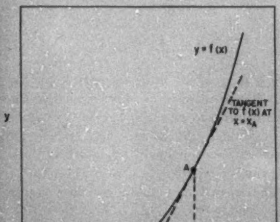


Fig. 5. Geometry of Newton's method

Aside from these simple equations, probably the greatest effort has gone into development of methods of solving higher degree polynomial equations. However, only two of the great variety of these will be mentioned. The first, which is equally applicable to other equation forms, is Newton's well-known method for real roots. It is assumed that the approximation x_A to the real root x_0 is known. The derivative of the function is evaluated at A, and the tangent to the function at A is projected to the x-axis, as in Fig. 5, to give the improved approximation x_B. The formal expression for this process is

$$\frac{y_A}{x_A-x_B}=\frac{dy_A}{dx}=f_A' \qquad (8)$$

or

$$x_B=x_A-\frac{f_A}{f_A'} \qquad (9)$$

The new approximation x_B can then be used in the same kind of expression to give a still better approximation x_C, etc. This process of using the result of a computation as the argument in a repetition of the very same computation is called iteration. Sometimes it leads to an answer (is convergent), and sometimes not. Newton's method is usually quite good in this respect, and inherently tends to converge rapidly. In most practical cases, the initial approximation can be reasonably rough, and may often be chosen on the basis of general knowledge of the physical situation.

It is worth while to realize, if the function cannot be differentiated, that Newton's method is only the limiting form of the method in which two nearby values of the function are calculated and the straight line joining these points used to find an estimate of the root. Both ways have their useful areas of application, and both can be used successfully when the function has more than one independent variable, or when several nonlinear functions of several variables must be solved simultaneously.

A particular case of two simultaneous nonlinear functions of two independent variables is that in which a given equation must be solved for its complex roots. The two independent variables are the real and the imaginary parts of the roots; the two functions are the real and the imaginary parts of the complex function. In the general case, their solution must usually be accomplished by the extended (2-variable) Newton's method, or multiple interpolation method. But in the case of complex roots of the general polynomial with real coefficients other ways have been developed, of which one is described as follows.

The basis of this method, which Aitken calls the "reduced penultimate remainder" method, is the determination of a quadratic factor of the complete polynomial expression. The polynomial with real coefficients has complex roots only in conjugate pairs, and each conjugate pair of complex roots can be expressed in the form of a quadratic factor with real coefficients; hence it is reasoned that the polynomial must have such quadratic factors in every case, and these are sought instead of the complex roots directly. Once a quadratic factor is determined, it may be solved by the quadratic formula. It is not universally agreed whether the real roots ought to be found and removed first, since the quadratic separation can equally well give pairs of real roots, too.

The method of reduced penultimate remainder is iterative, a first approximation to the quadratic factor being required. This approximate factor is divided into the polynomial by ordinary division or synthetic division, as preferred, the process being stopped before the last possible step of the division is performed. The remainder at this stage (the penultimate remainder) is also a quadratic expression, and is frequently a better approximation to a factor of the polynomial than the original trial divisor. It is therefore used as the next trial divisor, leading to an iterative process. As it leads to simplification of the arithmetic, the polynomial and the trial divisor are both written with leading coefficients of unity; accordingly, the penultimate remainder is "reduced" to that same form before becoming the new trial divisor. Thus the name, and this the process.

As an example, an attempt will be made to separate a quadratic factor of the equation

$$x^5+6x^4+11x^3+10x^2+6x+1=0$$

which is of the type met in the solution of linear control systems, except for the simple coefficients. There is no good rule for estimating a trial quadratic divisor, but one which works fairly often is to use the last three terms of the polynomial. Accordingly, the following factor will be tried:

$$x^2+0.6x+0.1$$

with results as follows:

$$\begin{array}{r|l}
 & x^3+5.4x^2+7.66x \\
\hline
x^2+0.6x+0.1 & x^5+6x^4+11x^3+10x^2+6x+1 \\
 & 1+0.6+0.1 \quad 0.54\ 0.77 \\
 & \qquad\quad 3.24 \quad 4.60 \\
\hline
 & \quad 5.4 \quad 7.66 \quad \mathit{4.86}\ \mathit{5.23} \quad \mathit{1}
\end{array}$$

The penultimate remainder is in italics. It gives for the next trial factor:

$$x^2+1.08x+0.21$$

It is easy to see how to continue the process. In the present case, it converges after four more trials to the factor

$$x^2+1.71x+0.35$$

and the quotient

$$x^3+4.29x^2+3.31x+2.84$$

each of which, equated to zero, gives possible solutions of the original equation. The quadratic is easily solved by the quadratic formula. It would probably be best in the case of the remaining cubic next to separate a real root; of course, this could have been done originally, as is sometimes recommended.

This method naturally does not always work so well as in the example given. Various modifications are available, with different convergence properties. Other completely different methods can also be used, including the 2-way Newton's method already suggested. These other methods involve more complicated procedures, but the Newton variety has the virtue of being applicable to nonpolynomial equations as well.

Simultaneous Linear Equations

In addition to appearing as the direct result of describing physical systems, as in the case of electric networks, simultaneous linear equations often arise in computation processes—as in curve fitting by substitution. They are also a result of the common practice of solving differential equations by means of finite difference approximation. This last source has two special features: It may lead to very large systems of equations, and, if the differential equation is nonlinear, the algebraic equations will also be nonlinear. Only two of the generally used methods of solving sets of simultaneous linear equations will be mentioned here but one of these, the iteration method, is applicable also to nonlinear systems.

SYSTEMATIC ELIMINATION OF VARIABLES

The method of solution by successive elimination of variables is the familiar process taught in high school and college. Several special circumstances arise when this process is utilized in automatic computation. The pattern of computation is usually fixed in order to make the procedure generally applicable; this sometimes leads to elimination in an unfavorable sequence or the occurrence of such difficulties as division by zero. The automatic process may not be able to recognize if the equations are not independent, and may supply wrong answers without warning of this condition. Where the number of equations is large, special arrangement of the computation may be required in order to avoid overloading the number storage capacity of the computer. To take care of such situations, many variations of the elimination method have been proposed.

The basic features are illustrated here by means of the following three equations:

$$\begin{cases} 8x+2y-3z=2 \\ 4x+10y+2z=3 \\ 2x-5y-9z=5 \end{cases}$$

In an automatic process, the first of these might be divided by 8, four times the resulting expression subtracted from the second equation, and twice the expression subtracted from the third, giving

$$\begin{cases} x+\frac{1}{4}y-\frac{3}{8}z=\frac{1}{4} \\ 0+9y+\frac{7}{2}z=2 \\ 0-\frac{11}{2}y-\frac{33}{4}z=\frac{9}{2} \end{cases}$$

The second of these expressions can now be divided by 9, one fourth of the result subtracted from the first equation, and 11 halves of it added to the third:

$$\begin{cases} x+0-\frac{17}{36}z=\frac{7}{36} \\ 0+y+\frac{7}{18}z=\frac{2}{9} \\ 0+0-\frac{55}{9}z=\frac{103}{18} \end{cases}$$

Finally, the third of these expressions is divided by $(-55/9)$ and the result used to eliminate z in the first two, in similar fashion:

$$\begin{cases} x+0+0=\frac{109}{440} \\ 0+y+0=\frac{129}{220} \\ 0+0+z=-\frac{103}{110} \end{cases}$$

Now, it is not claimed that this is the best possible way to solve equations, but it shows an interesting situation. All the steps of the solution were dictated by the left sides (coefficients) of the equations, whereas all the answers came from the right sides (constants). If these equations had to be re-solved several times with different sets of constants, all that would be necessary would be to remember the steps and apply them to each new set of constants to get the corresponding new set of answers. While this is possible, it is often more convenient to use a special way of "remembering" the necessary calculations, which will be illustrated by means of the same set of equations as before. Now, however, symbols will be used for the constants:

$$\begin{cases} 8x+2y-3z=a+0+0 \\ 4x+10y+2z=0+b+0 \\ 2x-5y-9z=0+0+c \end{cases}$$

The constant terms here have been purposely written in the same form as the variables came out in the previous answers. The steps now are exactly the same as before, and it will be observed that the constants a, b, and c remain in their original columns:

$$\begin{cases} x+\frac{1}{4}y-\frac{3}{8}z=\frac{1}{8}a+0+0 \\ 0+9y+\frac{7}{2}z=-\frac{1}{2}a+b+0 \\ 0-\frac{11}{2}y-\frac{33}{4}z=-\frac{1}{4}a+0+c \end{cases}$$

$$\begin{cases} x+0-\frac{17}{36}z=\frac{5}{36}a-\frac{1}{36}b+0 \\ 0+y+\frac{7}{18}z=-\frac{1}{18}a+\frac{1}{9}b+0 \\ 0+0-\frac{55}{9}z=-\frac{5}{9}a+\frac{11}{18}b+c \end{cases}$$

$$\begin{cases} x+0+0=\frac{2}{11}a-\frac{3}{40}b-\frac{17}{220}c \\ 0+y+0=-\frac{1}{11}a+\frac{3}{20}b+\frac{7}{110}c \\ 0+0+z=\frac{1}{11}a-\frac{1}{10}b-\frac{9}{55}c \end{cases}$$

The final set of expressions here can be used to compute in simple fashion the answers for any assumed set of constants; they constitute, in fact, the inverse, or general solution, of the original equations. It can be seen that they recall the steps of solution, not individually, but in a simple consolidated form. It is also clear that the work involved in obtaining this result is less than twice that of obtaining the original single solution. Cursory examination of the process indicates that in this presentation many quantities have been included unnecessarily. In automatic computation the letter symbols would not be stored explicitly, and there would be no point in storing the zeros, so each stage of the foregoing process would contain not 18, but only 9, numbers.

ITERATIVE SOLUTION

As with the elimination method, there are also many variations of the iterative type of solution. For simplicity, only one, generally known as the Gauss iteration, is discussed herein. The system of equations is examined and each of them associated with the variable which seems to be most strongly determined by it, a different variable with each equation. Each expression is then rearranged so that its associated variable stands alone on the left side. In the example, the most logical procedure is to use the variable with the largest coefficient as the associated variable, giving

$$\begin{cases} x=\frac{1}{8}\,(2-2y+3z) \\ y=\frac{1}{10}\,(3-4x-2z) \\ z=\frac{1}{9}\,(-5+2x-5y) \end{cases}$$

Assuming some initial estimate of the variables (often zero), these values are put into the right sides and new values computed. The new values then go into the right sides, and so on. The results of this process for several stages are as follows (the last column gives the true values):

$$\begin{cases} x: & 0.25 & -0.035 & -0.079 & -0.14 \\ y: & 0.30 & 0.31 & 0.45 & 0.48 \\ z: & -0.56 & -0.67 & -0.74 & -0.82 \end{cases}$$

$$\begin{array}{rrrr} -0.18 & -0.20 & \ldots & -0.25 \\ 0.52 & 0.54 & \ldots & 0.59 \\ -0.85 & -0.88 & \ldots & -0.94 \end{array}$$

That the process is not unique can be seen by the fact that another arrangement of the same equations is

$$\begin{cases} x=\frac{1}{10}(2+2x-2y+3z) \\ y=\frac{1}{10}\,(3-4x-2z) \\ z=\frac{1}{10}\,(-5+2x-5y+z) \end{cases}$$

Since the iterative process is a nonterminating one, which in some cases may not even converge, while the elimination method gives a theoretically accurate answer with a definite number of operations, the advantage of iteration is not too evident. It is, however, very handy for the type of equations encountered in the numerical solution of differential equations; these may be large sets of equations, each of which contains only a few of the variables. In hand computation, iteration has the additional advantage of automatically correcting any errors, and hence of permitting initial work to proceed with fewer figures than might ultimately be required. And, as already mentioned the principle is equally applicable to nonlinear systems.

LATENT ROOTS AND MODES

In a large class of problems one has to consider systems of homogeneous linear equations containing an unknown parameter in the coefficients of the variables. Such systems occur in the study of transients in electric networks or in feedback control systems, the parameter being the (complex) system frequency. In these cases, it is sometimes necessary to determine the frequencies (latent roots) which permit the equations to have nonzero solutions and, in addition, the nature of these solutions (modes). Such calculations fall under the general heading of linear equations, although some procedures require the frequencies to be found from solution of a polynomial or other nonlinear equation. The objectives of this paper permit no further treatment of this type of problem.

Matrix Algebra

The subject of linear equations can hardly be dismissed without mention of matrix algebra. The principles involved in the parts of this topic which are of primary interest in engineering computations are relatively easy to understand, and they are almost indispensable in trying to follow the many ingenious and fascinating methods which have been proposed. On the surface, the nature of matrix algebra is the definition and use of an operational notation which allows the description of a set of simultaneous equations in the form of three symbols: one for the set of coefficients of the variables, one for the set of variables themselves, and one for the set of constants. Because of the algebraic properties of these symbols, it is possible to use them in devising many procedures, both finite and iterative, which would be difficult or practically impossible to discover from the detailed equations.

Differential Equations

It has already been pointed out that a very popular way of obtaining solutions to differential equations at the present time is by making use of finite difference approximations to convert them into corresponding simultaneous algebraic equations. The latter are then solved by the methods which have been indicated, or similar methods; and it is hoped that the resulting solutions form an adequate estimate of the true solution of the differential equations—or, even better, of the physical problems. Such a happy result cannot be taken for granted, though, so there is still great opportunity for the exercise of good judgment and ingenuity in this process.

In this particular field it is worth emphasizing also that such methods are only a supplement to, and not a replacement of, the other available methods. For example, there are classical, closed-form solutions, which give compactly the influence of the parameters of the equations; variational (such as Rayleigh-Ritz) solutions, which permit successive improvements and sometimes error estimates, as well as giving good matching of the boundary conditions; and analog solutions, which are sometimes faster than numerical integration methods and sufficiently accurate for many engineering problems.

Error Analysis

The use of numerical methods leads inevitably to the question of accuracy of the results. This refers not only to the effects of errors or uncertainty of input data and of the mathematical formulation of the physical problem, which would affect the results of any type of analysis, but also to those errors which are inherent in the numerical processes. Aside from arithmetic errors, or blunders, which an effort is made to eliminate by suitable checks and tests, two important types of error are easily recognized. One of these is called truncation error. This is the type of error introduced by retaining only a finite number of terms of a series approximation (curve fitting); or by using finite increments to approximate derivatives and integrals (differential equations, which is still curve fitting); or by using an iterative process, which must necessarily be terminated after a finite number of repetitions. The other is called round-off error, which is probably self-explanatory; it comes from the necessity of retaining only a finite number of digits in any arithmetical calculation.

Much effort has gone into studies of these matters. Generally usable results are relatively scarce and usually difficult or impractical to apply. Good engineering judgment often offers the most useful means of estimating reliability. Which brings this paper right back to where it started, to the suggestion that numerical methods are often as much of an art as a science; that they are a fascinating study and a continual challenge.

Information Sources

The literature in the field of numerical methods is very extensive, both in the form of books and as articles in periodicals. The variety is also great, ranging from direct applications to specific problems, to abstract and difficult mathematical treatments. The following lists those books which have been found useful and most understandable from an engineering point of view; the list is not exhaustive, and it contains much duplication. Any of the books named offers a good entry into the field of numerical methods and leads directly, through its references, into the pertinent current literature.

References

1. RELAXATION METHODS, **D. N. deG. Allen.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954. Iterative solution of linear algebraic equations; approximation of ordinary and partial differential equations.

2. NUMERICAL METHODS, **A. D. Booth.** Academic Press, New York, N. Y., 1955. Basic principles of all the common methods.

3. LINEAR COMPUTATIONS, **P. S. Dwyer.** John Wiley & Sons, New York, N. Y., 1951. Primarily solution of linear equations by elimination methods; some matrix algebra; error analysis.

4. ELEMENTARY MATRICES, **R. A. Frazer, W. J. Duncan, A. R. Collar.** The Macmillan Company, New York, N. Y., 1947. Written from the point of view of practical applications, but might require some guidance for a beginner in this topic.

5. NUMERICAL METHODS OF ANALYSIS IN ENGINEERING, **L. E. Grinter** (editor). The Macmillan Company, 1949. Engineering applications of iterative processes, especially to structural and to field problems.

6. NUMERICAL ANALYSIS, **D. R. Hartree.** Oxford University Press, London, England, 1952. Introduction to all the usual methods; many comments on practical features and limitations.

7. INTRODUCTION TO NUMERICAL ANALYSIS, **F. B. Hildebrand.** McGraw-Hill Book Company, Inc., 1956. Comprehensive treatment of all usual topics except matrices and partial differential equations; extensive bibliography.

8. PRINCIPLES OF NUMERICAL ANALYSIS, **A. S. Householder.** McGraw-Hill Book Company, Inc., 1953. Mathematics of numerical methods, except solution of differential equations; extensive bibliography.

9. NUMERICAL ANALYSIS, **Z. Kopal.** John Wiley & Sons, Inc., 1955. Mathematics of polynomial curve fitting, differentiation, integration, solution of ordinary differential and integral equations.

10. NUMERICAL CALCULUS, **W. E. Milne.** Princeton University Press, Princeton, N. J., 1949. Practical features of all usual topics except matrices and partial differential equations; working rules with examples.

11. NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS, **W. E. Milne.** John Wiley & Sons, Inc., 1953. Complements reference 10 by treatment of matrices as well as differential equations; extensive bibliography.

12. METHODS IN NUMERICAL ANALYSIS, **K. L. Nielsen.** The Macmillan Company, 1956. Principles of all the usual topics.

13. NUMERICAL METHODS IN ENGINEERING, **M. G. Salvadori, M. L. Baron.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1952. Practical methods of solving equations, finite difference calculations, and solution of differential equations.

14. NUMERICAL MATHEMATICAL ANALYSIS, **J. B. Scarborough.** Johns Hopkins University Press, Baltimore, Md., third edition, 1955. Practical treatment of all the usual topics with comments and detailed examples.

15. RELAXATION METHODS, **F. S. Shaw.** Dover Publications, Inc., New York, N. Y., 1953. Iterative solution of linear algebraic equations; approximation of ordinary and partial differential equations.

16. RELAXATION METHODS IN ENGINEERING SCIENCE, **R. V. Southwell.** Oxford University Press, 1940. Iterative solution of frameworks, networks, and vibrating systems, including some ordinary differential equations.

17. RELAXATION METHODS IN THEORETICAL PHYSICS, **R. V. Southwell.** Oxford University Press, 1946. Iterative solution of field problems.

18. CALCULUS OF OBSERVATIONS, **E. T. Whittaker, G. Robinson.** D. Van Nostrand Company, Inc., Princeton, N. J., fourth edition, 1944. Finite differences, nonlinear equations, curve fitting, statistics.

19. TREATMENT OF EXPERIMENTAL DATA, **A. G. Worthing, J. Geffner.** John Wiley & Sons, Inc., 1943. Practical discussion of curve fitting and effects of experimental errors.

Vol.
Pt.
195

Electric Hygrometer

W. H. MEIKLEJOHN
MEMBER AIEE

AN ELECTRIC HYGROMETER is a device used to measure relative humidity (RH) by electrical means. Mechanical methods are exemplified by the hair hygrometer and the sling psychrometer. An electric hygrometer was developed by Dunmore[1] and used extensively in radiosondes to obtain upper-air data for the United States Meteorological Service. The electric hygrometer described in this report has a lower resistance than the Dunmore hygrometer. The lower resistance of this hygrometer allows direct operation of relays with the hygrometer current.

Operation

A schematic diagram of the electric hygrometer is shown in Fig. 1. The resistance of the hygrometer varies with the RH or water vapor pressure in the air. The pores of the glass contain a salt solution in the form of a film on the surface of the pore which absorbs more water when an increase in RH occurs. The resistance of the solution in the pore decreases and more current flows. Similarly, a decrease in current is obtained with a decrease in the RH.

Electrolytic conduction takes place between the porous metal electrodes. If the salt used has a decomposition potential lower than water, the salt will plate out, and the hygrometer will fail, because of loss of salt. Therefore a salt such as lithium chloride which has a decomposition potential higher than water is used, so that hydrogen and oxygen are plated out at each electrode. If these gases are allowed to accumulate the hygrometer will fail, because of the loss of water in the pores. To eliminate this difficulty, alternating current was used to cause recombination of the hydrogen and oxygen to form water.

Construction

The porous glass used in the hygrometer was obtained from the Corning Glass Works in the form of square sheets 2 inches on a side and in both 0.005- and 0.010-inch thicknesses; see Fig. 2. The porous glass occurs in a stage of the process of making Vycor brand 96% silica glasses, sometimes called reconstructed glasses. A soft alkali borosilicate glass containing many alkali compounds is molded into the desired shape, and then all the alkali compounds are leached out, leaving behind a porous network of silica with some of the leaching salts still in the pores. It is at this stage that the glass is used for the electric hygrometer.

The glass for the hygrometer was

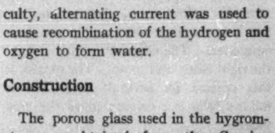


Fig. 1. Schematic diagram of electric hygrometer

Paper 58-8, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 10, 1957; made available for printing October 17, 1957.

W. H. MEIKLEJOHN is with the General Electric Research Laboratory, Schenectady, N. Y.

Fig. 2. Porous reconstructed glass plate used in the hygrometer

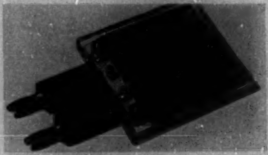


Fig. 3. Assembled hygrometer

treated to remove the leaching salts left in the glass during the leaching operation, and a 0.1 normal lithium chloride solution was introduced. The glass was placed in a breaker of distilled water for 3 days. The water was changed twice each day. The glass was then placed in a 0.1 normal solution of lithium chloride for 3 days. The solution was changed each day. Attempts to force the solution through the porous glass resulted in broken glass plates, even when the porous glass was backed with fritted glass which had been ground flat.

The surface of the porous glass was cleaned of salt by dipping the glass in distilled water several times. The porous glass was not allowed to stay in the distilled water long enough to change materially the concentration of lithium chloride in the pores.

The porous glass was placed in an evaporator and a 0.01μ (micron) film of gold and a 0.1μ strip of gold were evaporated on each side of the glass. The two strips were arranged perpendicularly to each other in order to reduce the area covered by such a thick film. The thick film would be less permeable to moisture and would cause an increase in response time of the instrument.

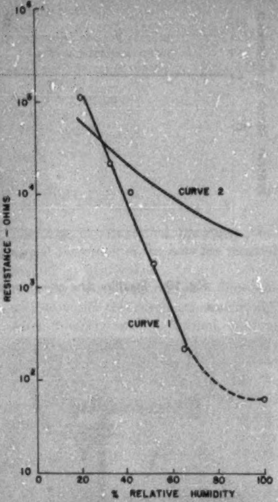


Fig. 4. Curve 1—Hygrometer resistance as function of RH. Curve 2—Calculated resistance variation

The porous glass was mounted between frames, as shown in Fig. 3. The glass was cemented to only one frame at only one spot. Copper wires of the proper diameter were used as spacers, so that the frame would not restrict the motion of the glass and cause cracking. The glass changes its dimensions considerably with changes in relative humidity as well as with temperature. Electrical contact was made to the evaporated gold strips by means of mechanical pressure.

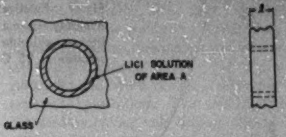


Fig. 5. One pore in reconstructed glass

Resistance

The variation of resistance of one hygrometer with RH is shown in curve 1, Fig. 4. The resistance change from zero to 100% RH is more than 4 decades. The resistance per square centimeter of porous glass surface in the humidity range of 20% to 65% is given by

$$R=8\times10^4e^{-0.188H} \quad (1)$$

where H = per cent of RH and R = ohms/cm² (per square centimeter). The resistance values are low enough to allow the hygrometer to be used in series with a relay to control some other device directly. The self-heat error in the hygrometer should be lower than the self-heat error of other electric hygrometers because of the lower resistance.

The expected resistance variation with RH may be calculated.[3] Consider one pore in the porous glass as shown in Fig. 5. The pore contains lithium chloride in solution of cross-sectional area A. The pore has a length L so that the resistance R is given by

$$R=\frac{L}{\sigma A} \quad (2)$$

where σ = conductivity mho/cm. There are the following relations:

$$\sigma=C\lambda \quad (3)$$

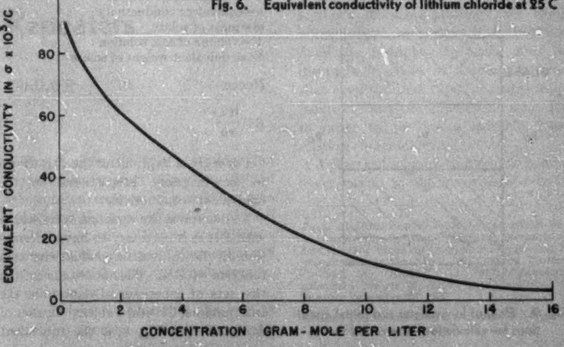


Fig. 6. Equivalent conductivity of lithium chloride at 25 C

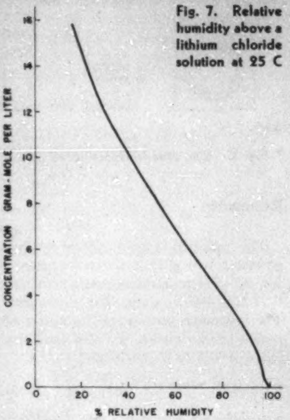


Fig. 7. Relative humidity above a lithium chloride solution at 25 C

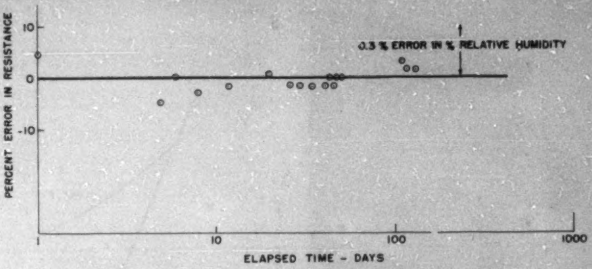


Fig. 10. Stability data on hygrometer having lithium chloride in its pores

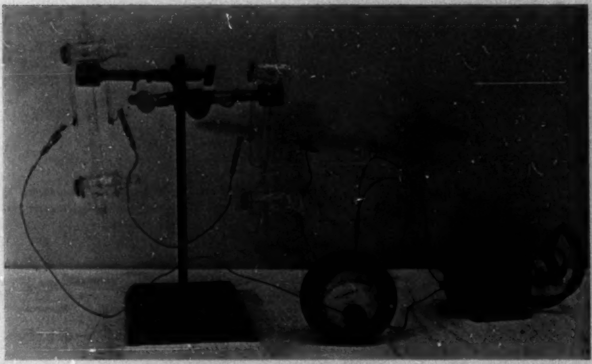


Fig. 11. Apparatus used to determine per-cent recombination of hydrogen and oxygen in pores of hygrometer

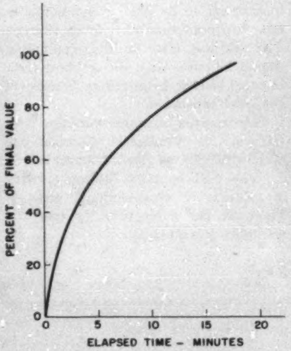


Fig. 8. Response time of reconstructed glass hygrometer

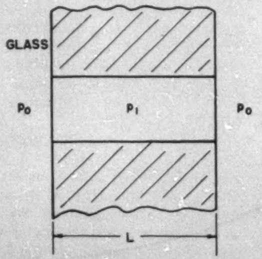


Fig. 9. Pore of hygrometer and initial conditions for calculation of response time

$$C=\frac{m}{VW} \tag{4}$$

where

C = concentration of salt
λ = equivalent conductivity
m = mass of solute
V = volume of salt solution
W = equivalent weight of solute

Hence

$$R=\frac{WL^2}{m\lambda} \tag{5}$$

For a given hygrometer the R is shown to be inversely proportional to the equivalent conductivity.

To determine the variation of resistance with RH it is necessary to have information on the equivalent conductivity as a function of RH. This is obtained from two sets of experimental data from the International Critical Tables. One set of data shown in Fig. 6, is the equivalent conductivity as a function of concentration. The other set, shown in Fig. 7, is the RH as a function of concentration. With these two sets of data and equation 5, the expected resistance variation has been calculated. The result is shown as curve 2 in Fig. 4. The agreement with experiment is only qualitative.

Response Time

The response time of the hygrometer to 63% of its final indication is about 7 minutes; see Fig. 8. The response time is long because the pores in the glass are so small. If the pores are made large enough (1μ) to be several times the mean free path of the water molecules in the air, then the response time should be less than one second. This has been verified experimentally, and some rough calculations will be given to show that a response time of one second should be expected.

The pores in the reconstructed glass are 40 to 60 Å (Angstrom units) in diameter and only 65% of the surface area is glass. If one assumes an equally spaced array of holes in the surface, then it is easily shown that the pore separation is about 70 A. However, the mean free path of water vapor molecules at standard temperature and pressure is about 600 A. It is difficult under these conditions for the water molecules to get into the pores, and correspondingly difficult for them to get out.

If the glass has a pore size of 10,000 A (1μ), then the water molecules can easily diffuse into and out of the pores. In particular, they will obey Fick's laws of diffusion.

Consider the 1-dimensional case of diffusion

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial X^2} \tag{6}$$

where p = partial pressure of water vapor.

This equation may be solved for the case shown in Fig. 9, where the initial condition is that of the hygrometer, removed from a jar having a vapor pressure p_1, and into the atmosphere at pressure p_0. The solution of this equation is

$$p=p_0+\sum_{n=0}^{\infty}\frac{4}{\pi}\frac{(p_1-p_0)}{(2n+1)}e^{\frac{-(2n+1)^2\pi^2}{4L^2}dt}\cos\frac{(2n+1)\pi X}{2L} \tag{7}$$

The infinite series converges so rapidly that the first two terms are a very good approximation:

$$p=p_0+\frac{4}{\pi}(p_1-p_0)\cos\frac{\pi X}{2L}e^{\frac{-\pi^2 dt}{4L^2}} \tag{8}$$

Assume that

p_1 = 8 millimeters of mercury: 50% RH at 20 C (degrees centigrade)
p_0 = 3 millimeters of mercury: 20% RH at 20 C
D = 1/4 cm²/second
L = 0.1 cm
T = 0.01 second

then

$$\frac{p-p_0}{p_0}=0.14 \tag{9}$$

Therefore, in 0.01 second, the vapor pressure at the center of the pore has reached 86% of its final value.

An ultrafine (1μ) porous glass filter was used to verify this result experimentally. After filtering a 0.1 normal lithium chloride solution through the porous glass, the excess glass was cut off. Gold was evaporated on each side of the glass and the response time was measured to be less than one second. The resistance was very high because of the greater thickness and the reduced total circumference of the pores.

The response time of the hygrometer can be greatly reduced by using porous glass with pore diameter in the range of 0.1 to 1μ.

Stability

Stability data on a hygrometer tested in a closed container above a salt solution at 52% RH and at a constant temperature of 25 C are shown in Fig. 10. Because of the high sensitivity of the hygrometer a 10% error in resistance at 52% RH results in only a 0.3% error in the indicated RH as shown in Fig. 10.

The principles involved in making this hygrometer stable are very simple. Since electrolytic conduction takes place in the hygrometer, there must not be any metallic ions present to plate out on the electrodes and hence change the conductivity of the solution. This condition is obtained by having only lithium chloride salt in the pores of the hygrometer. Since lithium chloride has a decomposition potential higher than water, hydrogen and oxygen are plated out at the electrodes. The gas formation at the electrodes would also result in instability, and therefore it is necessary to cause recombination of the gas. Hence, the hygrometer must be used on alternating current in order to have both oxygen and hydrogen available for recombination. If the frequency of the alternating current is high enough, the oxygen and hydrogen will recombine. The test shown in Fig. 11 was devised to determine whether 60 cycles per second is a high enough frequency.

The test consists of electrolyzing a 0.1 normal lithium chloride solution and collecting the gas. There was no gas formation observed. Assuming that the gas bubble to be detectable would have to be at least 0.1 cm in diameter, and that the liquid would first become saturated with the gas, one may calculate that the efficiency of recombination is at least 99.999%. It may be concluded that in a pure water vapor atmosphere this hygrometer should be very stable.

References

1. AN ELECTRIC HYGROMETER AND ITS APPLICATION TO RADIO-METEOROLOGY, F. W. Dunmore. *Bulletin*, American Meteorological Society, Boston, Mass., vol. 19, 1938, p. 225.

2. A SIMPLE THEORY OF THE ELECTRIC HYGROMETER, W. Schaffer. *Ibid.*, vol. 27, 1946, p. 147.

Transistorized Negative-Impedance Telephone Repeaters

ROBERT P. DIMMER
NONMEMBER AIEE

PRESENT switching plans which utilize automatic switching equipment in the telephone plant call for many changes in the layout of circuits and routing of traffic. Where yesterday traffic was handled in as direct a path as possible with a minimum of switching, today one or more different routings may be used to connect the same two telephones at different times.

The inception of this switching plan immediately presents a problem to the transmission engineer, for it is he who must keep the transmission loss of each individual circuit as low as possible in a connection in order that transmission will always be satisfactory on all calls irrespective of routing.

One might consider that the way to solve this problem would be to operate all circuits at zero or close to that loss, and thus make the transmission independent of routing. A study of the methods of accomplishing this would immediately show that it is not economically possible. In some cases the distances involved in a circuit indicate that even the best types of transmission facilities must be operated at some loss in order to provide a suitable transmission characteristic.

A practical solution would be to operate every circuit at the lowest loss possible

---

Paper 57-1016, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted March 26, 1956; made available for printing December 23, 1957.

ROBERT P. DIMMER is with the General Telephone Laboratories, Inc., Northlake, Ill.

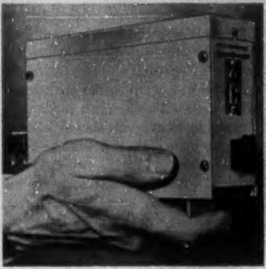


Fig. 1. AT-2 repeater

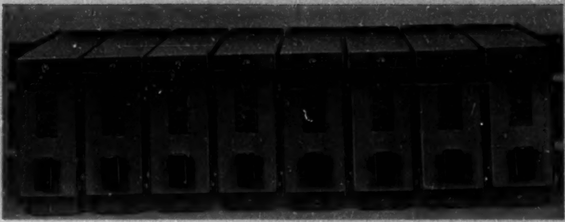


Fig. 3. AT-2 repeaters on shelf

considering its length and the type of facilities used. It would also be desirable to maintain these circuit losses within narrow limits because losses, either positive or negative, can accumulate on a connection with many links and produce either too low a speech level or excessive echo or hollowness.

One way to minimize the losses on 2-wire voice circuits would be to provide some type of voice amplifier or repeater to such circuits. In this case, would be necessary to treat each link of a connection in the same fashion, because a call might be routed through many different combinations of circuits at different times. In a typical telephone exchange there are many thousands of such circuits which might be used at one time or another to make the connections. Therefore, if this technique is to prove satisfactory, the voice repeater used should be one of small size and low power consumption, require practically no maintenance, and be low in initial cost. These features eliminate the hybrid type of repeater because of its size and cost, and suggest the negative-impedance type.

The negative-impedance type of repeater has been in use for some years and has proved its worth in many such applications. Unlike the hybrid-type repeater, this new repeater is essentially a device which, instead of amplifying in the usual sense of the word, in effect places a negative impedance in series with the line. The effect obtained with this repeater is to cancel an equal positive impedance naturally inherent in the line, and raise the transmission level accordingly. The shunt-type repeater was introduced some time after the negative-impedance type and eliminated some of the undesirable effects caused by the series unit. With both repeaters working as a team, the transmission engineer finally has a tool with which he can reduce the loss on multiswitched connections at a cost within reason.

The first negative-impedance repeaters employed vacuum tubes, but a forward step was taken when the newly developed transistor was considered for use in this unit. Transistors exhibit none of the practical limitations of the vacuum tube such as bulkiness, requirements of a heated filament, relatively short life, gradual deterioration with age, high anode voltage, or limited ruggedness. On the other hand, they do possess such desirable characteristics as small size, low power consumption, low voltage requirements, and apparent unlimited life. This paper will deal with a transistorized version of the negative-impedance-type repeater; specifically, the series-type *AT-2* and the shunt-type *AT-3* manufactured by the Automatic Electric Company.

Description

The series-type *AT-2* repeater is shown in Fig. 1. Its mate, the shunt-type *AT-3*, is identical in outward appearance. An inside view of the shunt unit is shown in Fig. 2 with an arrow pointing to the transistors. These transistors are mounted between two aluminum plates which act both as a mounting device as well as cooling fins to keep the transistors in a stable operating condition. A shelf of eight repeaters is shown in Fig. 3.

Either unit measures approximately 2 inches in width and $9^1/_4$ inches in over-all depth and a shelf of repeaters occupies a height of $5^1/_4$ inches. Eight repeaters will fill a single shelf designed to be mounted on a standard 19-inch relay rack, or ten units can be mounted on a 23-inch rack. Each repeater slides on a track to engage an 8-point shelf jack which is adjusted at the factory to cause the lines on both sides of the repeater to become connected when a repeater is removed. This feature permits rapid replacement of a defective unit if trouble should occur without interrupting an existing conversation.

Either repeater uses two conventional type-*2N44* alloy junction p-n-p transistors operating in a type of push-pull circuit. The adjustable part of either unit contains a number of graded values of resistance, capacitance, and inductance which may be combined by strapping in the field to give the desired network configuration and values in order that the repeater

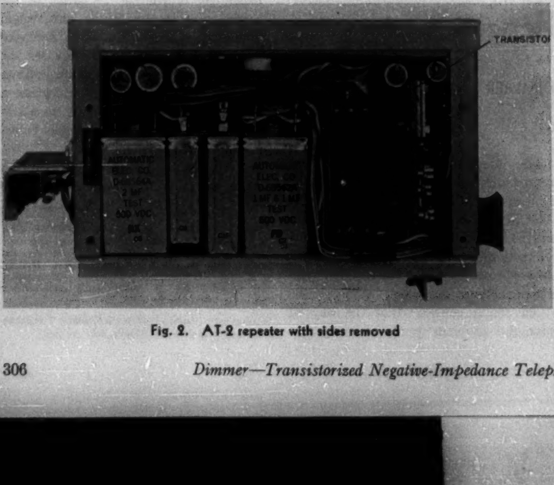


Fig. 2. AT-2 repeater with sides removed

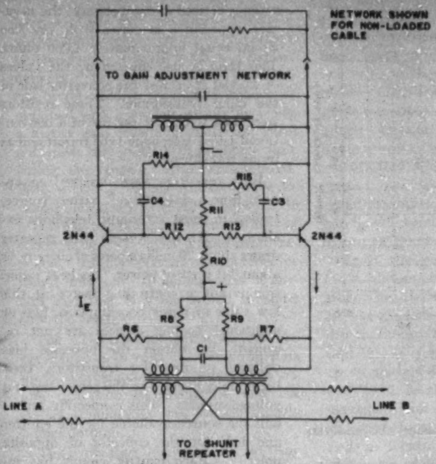

NETWORK SHOWN FOR NON-LOADED CABLE
TO GAIN ADJUSTMENT NETWORK
R14
R15
C4
C3
R11
R12
R13
2N44
2N44
R10
I_E
R8
R9
R6
R7
C1
LINE A
LINE B
TO SHUNT REPEATER

Fig. 4 (left). Circuit of series repeater AT-2

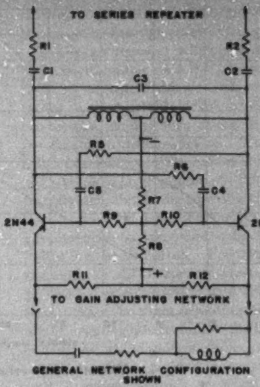

TO SERIES REPEATER
R1
R2
C1
C2
C3
R5
R6
C5
C4
R7
2N44
2N44
R9
R10
R8
R11
R12
TO GAIN ADJUSTING NETWORK
GENERAL NETWORK CONFIGURATION SHOWN

Fig. 5 (right). Circuit of shunt repeater AT-3

may be made adaptable to practically all exchange facilities. Access to the terminals connected to these components is provided by removing the top cover of either unit.

Strapping tables are furnished to enable the user to install these repeaters quickly on the most common types of cable. Where a mixture of cable is encountered, an installer's set is available which permits the rapid selection of network component values to satisfy the particular line condition. The values determined by this method would then be duplicated in the repeater strapping panel.

Operation

The negative-impedance-repeater concept has been covered by previous authors[1–5] on circuits using vacuum tubes; some few have considered the transistor approach. This type of repeater is different from the usual hybrid coil repeater, very familiar to transmission engineers, in that it uses only one amplifier and one network for a complete 2-way 2-wire voice repeater. The network in the new repeater performs the function of gain adjusting rather than balancing. By changing the constants in the network one can change the gain equally over the voice-frequency range, or if desired, the gain can be changed differently at various points within the band, thereby providing a means of equalization. No external filters or equalizers are required, and no additional signaling bypass equipment is needed. Direct current for talking battery and supervision and the various signaling voltages usually encountered in the telephone exchange pass through the repeater with very little difficulty. This type of repeater has the added advantage that, in the event of failure of some component within the unit, the circuit will still function but with its loss raised until the defective unit has been replaced.

The basic circuit of the transistorized series-type *AT-2* is shown in Fig. 4, and of the shunt-type *AT-3* is shown in Fig. 5. The repeaters are connected to a line by the method shown in Fig. 6. In this case the combination repeater would be called the *AT-23*. Either repeater consists of three principal parts: (1) coupling device; (2) balanced transistor amplifier or converter stage; and (3) network.

The series repeater is coupled into the line by a conventional transformer with a 1-to-1 ratio so connected as to insert equal amounts of negative impedance in series with each side of the line, to maintain line balance. This ratio was chosen to prevent high secondary transients from damaging the transistors. In the shunt repeater, the converter is coupled across the line through a resistance-capacitance network. The resistance provides the proper repeater gain, and

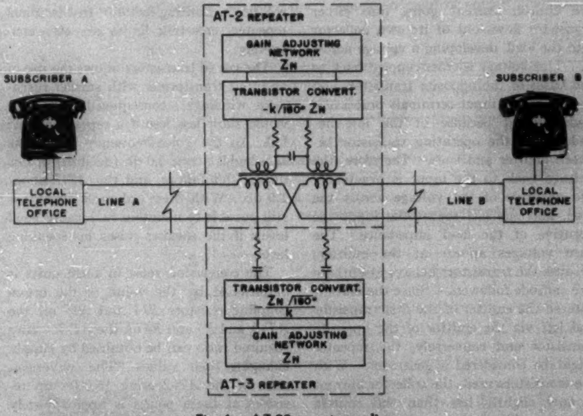

AT-2 REPEATER
GAIN ADJUSTING NETWORK
Z_N
TRANSISTOR CONVERT.
-k/180° Z_N
SUBSCRIBER A
SUBSCRIBER B
LOCAL TELEPHONE OFFICE
LINE A
LINE B
LOCAL TELEPHONE OFFICE
TRANSISTOR CONVERT.
-Z_N/180° / k
GAIN ADJUSTING NETWORK
Z_N
AT-3 REPEATER

Fig. 6. AT-23 repeater on line

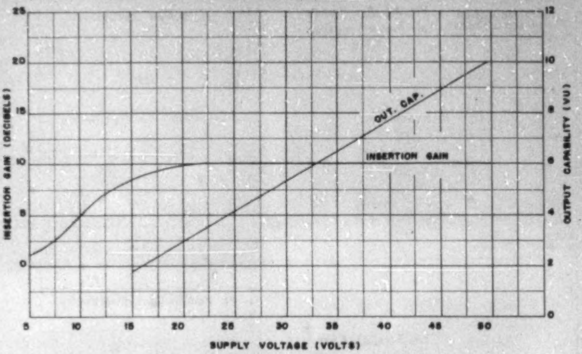


Fig. 7. Effect of supply voltage change on AT-2 and AT-3

the capacitors serve as isolation for any type of d-c signaling or dialing on the telephone circuit. It can be seen that the shunt repeater circuit is basically the series circuit in reverse; in this case the collectors of the transistors serve both as input and output terminals of the device, and the gain-adjusting network is connected to the emitters.

To analyze the operation of the circuit in Fig. 4, assume that two subscribers are engaged in a conversation on a particular telephone circuit which contains a series-type transistorized negative-impedance repeater similar to that shown in Fig. 4. Their voice signal is coupled to the emitters of the transistors in a push-pull fashion. Because of the characteristics of p-n-p junction transistors with an alpha value close to 1.0, practically all the emitter current going into either transistor flows out of its own collector into the load, developing a voltage across it. This voltage is cross-coupled back to the bases of the opposite transistor, and thence to the input terminals practically undiminished because of the low impedance of the operating transistor between emitter and base. Therefore, the voltage back to the input is practically the negative of the voltage across the load, and one can see at the input the negative of the load impedance. The base voltages appear at the emitters because the transistors behave essentially like cathode followers. Since the current entered the emitter of the first transistor and left via the emitter of the opposite transistor and conversely, the repeater might be considered a generator. With the transistors used, the collector current is only slightly less than the emitter current, and the emitter-to-base resistance is under 100 ohms; hence this analysis seems quite correct.

As is the case with all repeaters of the series type, the device is short-circuit unstable. In terminal repeater applications, the series repeater will oscillate if a short circuit is encountered on the line side during a dialing condition. The shunt repeater has the opposite characteristic, that is, it is open-circuit unstable. If the line should become open-circuited on the side near the repeater, the unit will oscillate during an idle circuit condition, assuming that no idle circuit termination is used. Generally speaking, a dialing condition for the series repeater and an open-idle-circuit condition for the shunt repeater are the controlling factors governing the amount of gain possible. Oscillation which occurs due to these conditions will be the limiting factor. In a terminal repeater, crosstalk limits are also controlling.

The use of transistors allows the use of an input transformer with smaller resistance windings; consequently, the line would show less loss if a repeater should fail. In the voice-frequency range, the *AT-2* adds about 6.0 db (decibels) of loss during this failure, and the *AT-3* about 1.0 db. With these orders of losses it is still possible to maintain reasonably good levels if the speaker raises his speaking level.

The conversion ratio in these units is determined by the value of the cross-coupling resistors *R14* and *R15* on the *AT-2*, and *R5* and *R6* on the *AT-3*. Any required ratio can be obtained by simply changing their values. The conversion ratio of the *AT-2* using 15,000-ohm resistors at these points is approximately 9 to 1. For example, if a 9,000-ohm resistor is used as the network, the negative impedance in series with the line would equal approximately 1,000 ohms. This ratio is also effected by the values of *R6* and *R7* across the converter side of the input transformer. These resistors were used to permit the use of a conventional 1-to-1 telephone-type repeat coil as input transformer.

Both repeaters were designed to operate from a single 48-volt battery source, typical in most automatic telephone exchanges. At this voltage each repeater draws about 20 milliamperes of current, or about 1.0 watt of power. As heat radiation is not a controlling factor at this low value of power consumption, fans or additional cooling devices are not required. To obtain the opposite bias voltages required by transistors, their bases are connected to the center tap of a voltage divider. This connection makes half the source available for the emitter and half for the collector of opposite polarity. Each emitter circuit has individual dropping resistors to set their operating point. This method has proved more satisfactory when transistors with wide variations are encountered. It also permits the control of repeater gain below 500 cycles by changing the value of *C1*. This limitation offers some desirable features such as greater gain in the voice band, a certain degree of noise suppression, and the reduction of the series repeater's effect on various signaling voltages. It also affords a control of the low-frequency oscillation which this repeater originally developed on nonloaded telephone circuits.

The conversion property or alpha of a transistor varies little with bias conditions, and consequently a change in supply voltage should have little effect on the insertion gain of a series repeater within certain limits. Fig. 7 shows the change that occurs in the gain of this unit when the supply voltage is changed from 0 to 50 volts. Note that the gain remains practically constant from 20 to 50 volts. This feature permits the pole-mounting of repeaters and supplying them with power from a distant office by using a spare pair in the cable. It might be possible, in some cases to supply each repeater over its own line pair, but each installation would have to be engineered differently in view of the many types of signaling and dialing one encounters in telephone systems. The drop in supply voltage affects only the output capabilities of a series repeater. As this voltage drops as it is extended farther and farther from the office, so also does the input signal; consequently, one would still have a satisfactory repeater. The gain of the shunt repeater drops in proportion to the supply voltage with a constant network. The gain of this unit however, can be returned to the standard value as obtained with a 48-volt supply by using a higher gain network. This unit does not have any output capabilities limit within the range of levels encountered. With a 24-volt supply source, and with proper selection of the *AT-3* network, it is possible to realize full output with the combination *AT-23* as a mid-point repeater. In this application each repeater would draw only 10 milliamperes of current and thus the pair would consume only 0.5 watt.

The *AT-2* repeater with a supply voltage source of 48 volts will handle outputs up to approximately 10 volume units before distortion is perceptible, and the *AT-23* combination will handle slightly higher outputs.

The repeaters have remained undamaged up to a temperature near 100 C (degrees centigrade) although at this high temperature other components within the units did become damaged. The gain of the unit remained constant up to about 60 C, as shown in Fig. 8. Beyond this temperature the gain dropped until it reached zero at about 90 C. Compensation was considered, but temperatures in telephone exchanges rarely reach these values. For pole-mounted repeaters a fan should be used when temperatures above 60 C are reached. If these high temperatures are reached at any time, the transistor invariably returns to normal when the temperature is reduced to normal.

The transistor in both units undergoes an aging process during the first few hours of operation. This process causes the beta of the devices to drop a few per cent. This reduction in gain happens on both repeaters, but is more noticeable on the *AT-3* because of its high impedance output. In all cases, the units stabilize at the end of 24 hours. At present, the transistors are stabilized before being connected into the units by being baked for a period of time at 85 C.

The operation of the repeaters under high transient conditions was further considered. To simulate a high transient, a large capacitor, charged to a potential of more than 1,000 volts, was discharged across the input terminals of the *AT-2* repeater. This unit showed no signs of damage after over 1,000 discharges, whereas a comparable vacuum-tube unit would be inoperative after 10 to 20 discharges.

It might be possible to use an n-p-n and a p-n-p transistor in a complementary symmetry circuit with some simplification; however, matched transistors of the two types with similar operating characteristics might be difficult to obtain. A single-transistor negative-impedance repeater has been successfully tested, but like vacuum-tube circuits with one tube versus those with two tubes in push-pull, greater care was required to prevent signal distortion, and the unit could not handle sufficient output levels before this distortion point was reached.

Under the most favorable conditions the maximum integrated gain of an *AT-2* alone can never exceed approximately 50% of the loss of the line with 10 db as the maximum limit, and the combination *AT-23* repeater can never exceed approximately 85% of the loss of the line with 14 db as the maximum limit. The maximum gain depends upon the length or loss of the trunk or line, its uniformity and the location of the repeater. It is a function of the negative impedance of the series unit, the negative admittance of the shunt unit, and the sums of the impedances of the two lines looking away from the repeater between working and idle circuit conditions. The maximum gain when an *AT-2* alone is used on a loaded line is approximately 8 db. As a terminal repeater the repeater gains are reduced about 20% with a crosstalk limit of approximately 8 db as high. The integrated gain is the insertion gain taken at several points over the voice-frequency range and divided by the number of points; it gives a more accurate picture of repeater performance. A given gain versus frequency characteristics may be reshaped at the lower or upper frequencies by only a minor adjustment of one of the network components without significantly changing the average value of the gain.

The series repeater will perform very satisfactorily alone on a line which is not too long or where the amount of delay (echo) is short. This delay, or echo, is produced when any such device is placed in a smooth transmission line. When the delay is small, as in the case of an interoffice trunk, a large amount of "echo" can be tolerated, but in the case of a toll connection the total delay of many connected circuits may be large and only a small "echo" can be permitted. For these cases the shunt or *AT-3* repeater was designed to be used in conjunction with the series unit. Since the shunt repeater would add shunt negative admittance to the line and the series would add a series negative admittance to the line and the series would add a series negative impedance, the combination would least tend to change the over-all characteristic of the line. The image impedance of the repeater in this case would more nearly match the characteristic impedance of the line. Such a combination repeater would provide approximately the same gain as the hybrid repeater under similar line conditions, and have similar effects on return loss. In very few cases in the local telephone plant would any gain higher than the 14 db possible with an *AT-23* be required.

At the present time about three *AT-2* repeaters are being used to every *AT-23* used. However, it is safe to assume that in the next few years the combination repeater will be employed in the majority of cases.

Because the *AT-2* is in series with the line, this repeater has, in the conventional sense, no input or output impedance and

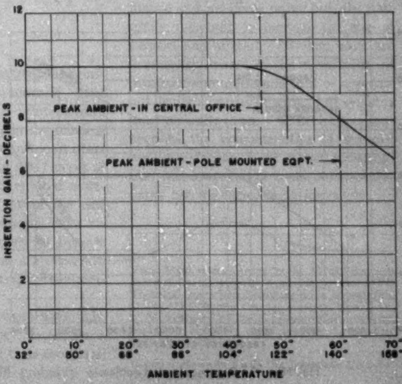
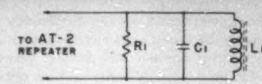
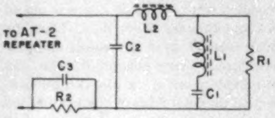


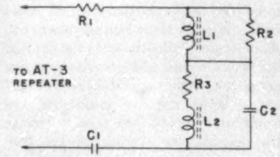
Fig. 8. Effect of ambient temperature change on AT-2



AT-2 NON-LOADED CABLE NETWORK (A)



AT-2 LOADED CABLE NETWORK (B)



AT-3 LOADED CABLE NETWORK (C)

Fig. 9. Networks used on repeaters

cannot be correctly tested unless connected to the line in question or a closely simulated condition. The shunt unit has a high output impedance in the vicinity of 8,000 ohms at 1,000 cycles with a resonant frequency of 130 cycles. For those few circuits using 135-cycle signaling the use of an *AT-3* is not recommended. Below 100 cycles the shunt repeater does not load the circuit to any extent, and at 20 cycles the effect is negligible.

The stability of both repeaters against oscillation is obtained by setting the network for gain values which will be stable under trunk or line conditions, and by selecting network values that drop the gain at the low and high end of the band. These units will invariably be stable under trunk or line conditions, if they are stable under idle conditions.

Application

In the 3 years that these repeaters have been in service, they have found a wide field of application in many types of circuits. A few of the more widely used of these applications are as follows:

1. On special service trunks, private branch exchange extensions, and tie trunks.
2. On longer tandem trunks to reduce over-all tandem standards.
3. On tributary trunks where extensive cable is used.
4. On the longer toll connecting trunks.
5. On trunks to centralize intercept and information desks.
6. To equalize and add gain on nonloaded cable in the same manner as is done with the use of loading without touching the cable proper.
7. In some cases on long subscriber lines where cable is used.

Performance

When *AT-2* repeaters are placed in service, two general types of networks are used depending on the types of lines connected to both sides of the repeater, and whatever central office equipment might be connected across the lines at the repeater. Fig. 9 shows these networks. The network in Fig. 9(A) is used when an open wire or nonloaded cable is connected to both sides of the repeater, or an open wire is connected on one side and a nonloaded cable on the other. The resistance *R* and the capacitance *C* determine the approximate over-all gain and frequency equalization. It is possible to change the frequency response of the repeater by selecting various values of *R* and *C*. If some type of long-line adapter, repeating coil, or comparable equipment is located near the repeater, an additional inductance is connected across the network. The performance of a series repeater in an application of this type is shown in Fig. 10. This curve shows how a nonloaded line with its greater loss at the higher frequency end can be equalized directly at the repeater to produce a circuit which, in effect, is similar to that obtained if the line were loaded.

When an *AT-2* is installed at the midpoint of a loaded line, as a terminal repeater at the end of a loaded line, or as an intermediate repeater between a loaded and a nonloaded line, the network shown in Fig. 9(B) is used. Capacitor *C2*, commonly referred to as the building-out capacitor, is determined by the end sections of the loaded line facing the repeater. The inductance *L1* and *L2* and capacitor *C1* comprise a low-pass filter with *R1* as one side termination. This filter determines the manner in which the gain is effected near the cutoff frequency of the line. In cases where a loaded line is connected to one side of the repeater and a nonloaded line to the other side, an additional capacitance *C3* and resistance *R2* are added to handle the nonloaded line characteristics.

An application of this type is shown in

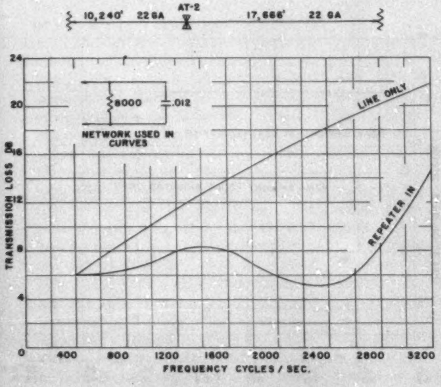


Fig. 10. AT-2 on nonloaded cables

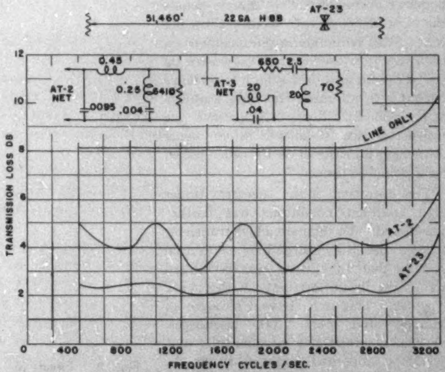


Fig. 11. AT-23 on loaded cable

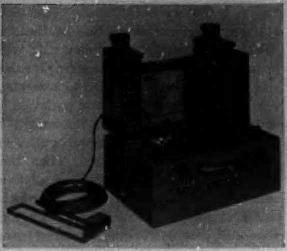


Fig. 12. Installer's set

Fig. 11. Note that the response curve of the series repeater is a series of peaks, which are further accentuated when telephones are used as terminations at each end instead of resistances. These result from variations in line and telephone impedances and make it rather difficult to judge accurately the effective gain of the repeater. When a repeater is properly installed, these variations do not exceed ±2.0 db.

When a combination repeater is required the gain adjustment network of the shunt unit would have a configuration as shown in Fig. 9(C). The capacitance *C1* governs the amount of low-frequency gain, and the inductance *L1*, governed by its shunt resistance, governs the amount of high-frequency gain. The resistance *R1* governs the over-all gain. If the *AT-3* is to be used as a terminal repeater, an additional resonant circuit may be added in series with the network to reduce a peak which sometimes occurs under these conditions. The shunt repeater adds a small amount of gain to the *AT-2* but, more important, it tends to straighten out the response curve.

Installation

It is readily apparent from the previous text that the most difficult part of the installation of a repeater of this type is in selecting a proper gain-adjusting network. When this is found the repeater will require no further attention other than an annual check of the gain of the unit, assuming that the lines have not been altered. This network can be determined by a cut-and-try method with reasonable accuracy; however, to speed up this selection an installer's set was designed to permit one to adjust these various components in the network quickly and accurately. A view of the installer's set with a series and shunt repeater attached, is shown in Fig. 12. By using the necessary oscillator and voltmeter one man can make all adjustments directly at the repeater and measure the final insertion gain of the unit. The use of this set simplifies the installation procedure and enables one person to obtain a satisfactory network in a short time. If a number of repeaters are to be installed on other pairs in the same cable, it is only necessary to determine the network with one repeater, and duplicate the strapping in the other units.

Conclusions

Today, several thousand of these repeaters are giving satisfactory service Some of these units have been in operation continuously for 3 years on their initial transistors. With aged transistors and close tolerance components, present repeaters are quite uniform in gain characteristics.

With small-size and low-power repeaters such as these available, it should be economical to apply them and provide quality transmission on finer gage cable than heretofore was possible. As transmission standards are raised, there may be cases where a still smaller gage trunk with two repeaters will be cheaper than a larger gage circuit with one repeater.

References

1. TRANSISTOR NEGATIVE-IMPEDANCE CONVERTERS, J. G. Linville. *Proceedings*, Institute of Radio Engineers, New York, N. Y., June 1953.

2. A NEGATIVE IMPEDANCE REPEATER, J. L. Merrill, Jr. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1461–66.

3. NEGATIVE-IMPEDANCE TELEPHONE REPEATERS—APPLICATION IN THE BELL SYSTEM, Arthur F. Rose. *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 430–35.

4. THEORY OF E-TYPE REPEATERS, J. L. Merrill, Jr. *Ibid.*, pp. 443–47.

5. E1 TELEPHONE REPEATER. *Instruction Bulletin no. 1274*, Western Electric Company, New York, N. Y., 1950.

The MITE Teleprinter

BERNARD HOWARD
NONMEMBER AIEE

Synopsis: This newly developed telegraph page printer weighs only 12 pounds, but is completely compatible with existing teleprinter equipment. It is believed to be the first practical teleprinter of radically reduced weight and cube and one of the most radical innovations in the teleprinter art. This paper describes some characteristics of the machine and discusses the basic mechanisms used.

A NUMBER of new "data read-out" devices have been developed in recent years to keep pace with the requirements of automatic digital computers. The outputs have taken the forms of cathode-ray indicators, magnetic tape, punched cards and tape, smoke printing, magnetic printing, and electrostatic printing, as well as alpha-numeric page printers.[1] Page and tape printers have been used for many years in communication circuits and have been made to operate in response to as many as 11 types of Morse, synchronous printer, and start-stop printer codes.[2] Many of the new types of equipment, however, have required special input signals or operate at high rates of speed and are otherwise not compatible with existing communication equipment.

Civilian and military users have indicated that, although they anticipate a need for higher speeds in their communication circuits, they will probably use printers operating from the telegraph start-stop code for some time to come.[3,4] The Armed Forces find it increasingly desirable to add electronic aids to their mobile and air-borne units but are naturally reluctant to increase the size of vehicles and aircraft. Civilian users are increasingly aware of the desirability of printed communications and of a print-out device compatible with those systems most generally in use. In both these connections the characteristics of reduced weight and cube, lower power consumption, and lower initial costs and maintenance become increasingly important.

Paper 58-48, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted September 30, 1957; made available for printing December 3, 1957.

BERNARD HOWARD is with the Teleprinter Corporation, Paramus, N. J.

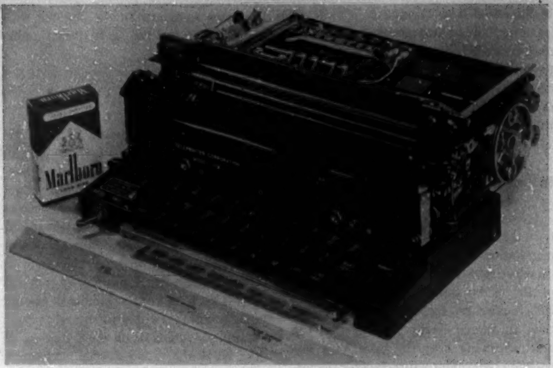


Fig. 1. MITE with keyboard in operating position

General Characteristics

It was with these facts in mind that the "Miniaturized Integrated Telegraph Equipment" (MITE) was developed. The initial machines were designed with a view towards both commercial and military specifications. It is believed that this small, lightweight printer has a place both in communication services and in certain computer applications. It is not intended to compete with high-speed units such as a recently announced digital printer which prints 10,800 lines (the equivalent of an average book) in 1 minute. The MITE is a page printer which weighs 9 pounds when receiving only and 12 pounds when fitted with a keyboard for sending as well as receiving. The printer unit is $3^{1}/_{2}$ by 8 inches by 12 inches. The keyboard unit is $1^{1}/_{2}$ by 8 by 12 inches. The carrying case is 5 by 11 by $12^{1}/_{2}$ inches. The carrying case weighs approximately 4 pounds, so that the complete equipment weighs 16 pounds. Fig. 1 shows the MITE with the keyboard can be pushed back under the printer, or it can be removed altogether as shown in Fig. 2.

The machine operates up to 100 wpm (words per minute) with the standard 7.42-unit Baudot code but future models are expected to operate up to 200 wpm if the code pulses are fed to the machine simultaneously. The machine is completely compatible with 7.42-unit code machines and can be converted in a matter of minutes to 60-, 66-, 75-, or 100-wpm operation by changing one externally mounted gear. Actually there is nothing inherent in the design of the machine which limits it to this type of code and special machines are being built to operate with other codes. Both the keyboard and type cylinder can be changed to accommodate weather symbols, or, in fact, to any desired selection of characters or functions up to a maximum of 64. (Two of these are shift from lower to upper case and the reverse shift from upper to lower.)

A New Approach to Character Positioning

One of the design objectives was to produce a machine capable of high-speed operation with a minimum of high-speed mechanical motions. This demanded a radically new approach to the method of deriving 64 mechanical positions from the five intelligence pulses in the start-stop code.

One of the basic principles used in the MITE is illustrated in Fig. 3. This uses a unique application of one of the six "simple machines," the pulley. It will be seen that if one end of the string is fixed and the other end is attached to a weight, a 1/2-inch movement of any pulley in a vertical direction will move the weight 1 inch. Next, suppose that each pulley is moved a different distance. Then the weight can be moved to any one of five different positions, depending on which pulley is moved. Finally, consider the possibility that each pulley is moved from its resting position twice the distance that the previous pulley travels and consider, also, that from one to five pulleys are moved at the same time. It can now be seen that there are 32 evenly spaced positions of the weight for all of the possible combinations of pulleys moved or not moved.

If a strip of type is substituted for the weight, then an elementary character-positioning device will have been provided. The 32-character length of the strip would, however, require a rapid motion of the type strip if the first and last letters on the strip were called for sequentially.

The space and motion requirement could be reduced by arranging the letters in the form of a rectangle eight characters long and four characters high and employing two co-ordinate motions. The space and motion requirement could be still further reduced by folding the rectangle into a half-cylinder eight characters long and with four faces and converting the linear co-ordinate motion used to choose one of the four up and down positions of the rectangle to rotary motions up to 180 degrees. This has been done in the MITE with an 8-sided type cylinder which is eight characters long; see Fig. 4.

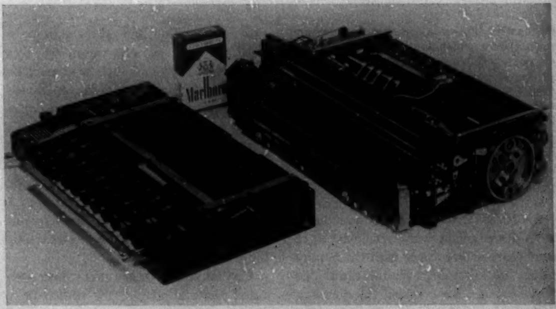


Fig. 2. MITE with keyboard removed

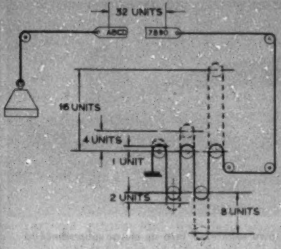


Fig. 3. Schematic of a simplified print-positioning device

The 32-position half-cylinder derived from the 5-pulley system is represented by four faces of the 8-sided cylinder and eight lateral positions along those faces. The opposite four faces of the cylinder and the eight lateral positions along those faces represent the upper case of the 32 basic positions so that the cylinder is rotated 180 degrees by the selection of letters or figures (lower or upper case) and the selection within that 180-degree segment is achieved by the basic 5-pulley system. Since a space is designated on the cylinder for every character or function that can be selected, the cylinder provides for 64 possible characters and functions, two of which are letters and figures. In this way a small cylinder can accommodate 64 characters of standard teleprinter type. In operation the first two of the five intelligence code pulses control the rotary selection motions of the type cylinder while the last three pulses control the lateral motions.

The Selector

The problem of arranging for the proper pulley to respond to the appropriate signal pulse is met by the selector in conjunction with motor-driven camshafts.

SYNCHRONIZATION

Synchronization of the MITE printer with the transmitted signal, as with other 7.42-unit code printers, makes use of two pulses. These are referred to as the "start" and "rest" or "stop" pulses. The start pulse effectively starts the rotation of the timing camshaft by releasing the start clutch. The rest pulse, which is 1.42 units long, occurs between the fifth signal pulse and that start of the next start pulse and allows a margin of error to exist between the motor speeds in the sending and receiving machines. The receiving camshaft rotates faster than the corresponding transmitting camshaft and comes to a stop during the stop pulse while the transmitting camshaft will rotate continuously during a continuous automatic transmission such as from a tape transmitter.

REDUCTION OF MECHANICAL SPEEDS AND LOAD

In the MITE, the mechanism which controls the pulleys, and consequently the type cylinder, is arranged to reduce the speed of operation and mechanical load on the machine by gradually positioning the type cylinder while the signal pulses are being received rather than by storing the pulses mechanically as they are received and beginning the positioning of the type after the receipt of the last intelligence pulse. For this reason, and because the type cylinder is located behind the printing medium, there is no need for a neutral position of the type cylinder; and consequently, in the case of a repeated character, the pulleys and cylinder do not move.

SELECTION

Each of the five type-positioning pulleys described in the foregoing is positioned by a cam follower passing through the pulley housing. The cams which move these followers are each coupled to a clutch. The five clutches and their cams are mounted on a continuously driving shaft on which the clutches free-wheel when not engaged and to which they are capable of being sequentially coupled.

In response to its signal as received by the magnetic selector, each clutch will either engage the driving shaft for a 180-degree rotation or hold its position. The action of the clutch in each case depends on whether a mark or space signal is received and whether the clutch was left in mark or space position by the previous selection. If, for example, the no. 1 pulse for a given selection is a mark pulse, and if the no. 1 pulse of the preceding selection were a mark pulse, the no. 1 clutch will remain stationary. If a spacing pulse is received for no. 1 and the preceding no. 1 was a mark, the clutch will engage for a 180 degree rotation turning the cam affixed to the clutch correspondingly. The cam follower and the pulley controlled by it will go the second of their two possible positions.

The actual release of each clutch is controlled by a pair of levers, one of which will engage the clutch and the other of which will stop it after 180 degrees. Each pair of levers is controlled by the selector and the instant of their release is determined by a sequence shaft which starts rotating upon receipt of the start pulse. The control of the clutch releases by the sharp drops in the sequence shaft cams is a feature which contributes markedly to accurate response and good range in the receiver.

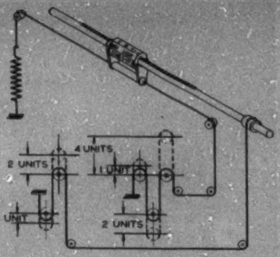


Fig. 4. Schematic of print-cylinder positioning method used in the MITE

At the end of the receiving cycle the sequence camshaft rests while the main driving shaft on which the five pulley clutches now free-wheel continues to turn. The rotation of the main shaft after the receipt of the fifth intelligence pulse releases the special function causing mechanism and either performs a mechanical function when one is selected or releases the print hammer to complete the character cycle.

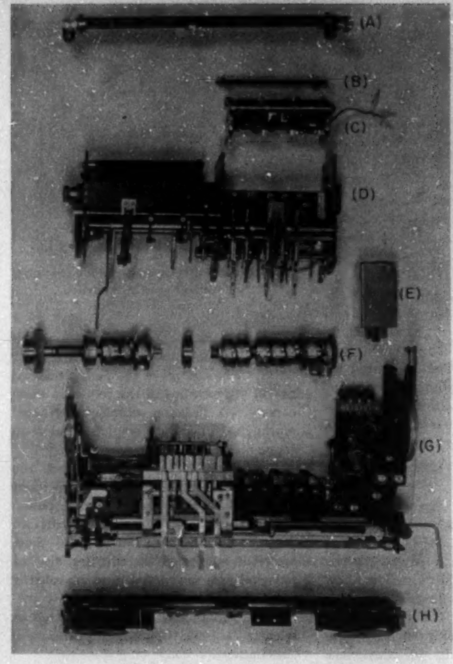
RANGE ADJUSTMENT

The relative time at which the magnet armature position is sensed can be adjusted by moving the angular position of the timing camshaft with respect to the point at which the camshaft is released by the start clutch. This angular position of the start clutch release mechanism is connected to a calibrated dial which is marked in terms of percentage of one signal pulse. The variation in angular position which is possible without mutilation of a message is known as the "range" of a teleprinter. The MITE has regularly been observed operating with ranges in excess of 75 points at 60 wpm in the laboratory, and has tolerated signal distortions in excess of 35 points under the same conditions. This range adjustment permits the printer to operate correctly from distorted signals such as may be received over long telegraph lines.

Keyboard

GENERAL

The mechanical and electric connections between the keyboard and the printer are very simple. The keyboard



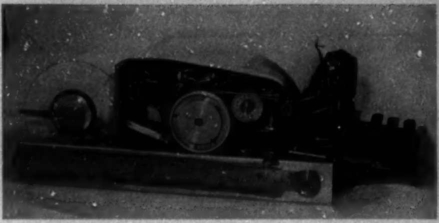


Fig. 5 (left). Printer shown separated into its major subassemblies

A—Rear plate
B—Selector clamp
C—Selector
D—Rear frame assembly
E—Transistorized switcher
F—Main shaft
G—Printer frame assembly
H—Ribbon magazine

Fig. 6 (above). Printer, keyboard, and paper roll shown mounted on a chassis

circuit is in series with the printer signal circuit and when the keyboard is removed from the printer or put in storage position, the contacts on the printer are automatically short-circuited. A spur gear couples the printer motor to the keyboard operating shaft. The keyboard can be operated separately as a unit remote from the printer by use of a separate motor. The keyboard operates as follows: The striking of any key, or the space bar, releases a clutch which permits the motor-driven keyboard camshaft to make half a revolution. At the same time five code bars are set up in a particular sequence according to which key is struck. These code bars will permit or block the motion of a series of cam follower levers, which in turn will permit a series of five contacts to make contact or stay open and thus transmit the proper signal. The contacts are in series with the signal line and when no keyboard key is pressed, the contacts are closed.

REGULATION OF TRANSMISSION

In order to transmit the correct signal, these contacts must make and break at particular times, which would normally require a very accurate adjustment of each contact. This is avoided by providing a master pulsing cam on the keyboard camshaft. This operates a pair of contacts which accurately determine the timing of transmission while the five contacts affected by the code bars only determine whether or not each of the pulses is transmitted. The stop on no. 2 and no. 4 contacts are connected to one of the pulsing contacts and the no. 1, 3, and 5 contacts are connected to the other pulsing contact. In this way the adjustment of the five contacts is made noncritical and only the single pair of master pulsing contacts requires adjustment.

Motors

The rotational inertia of the motor serves to integrate the pulses of mechanical power required for the printing cylinder positioning operations over the full character cycle. As the total power required by the machine is 25 watts including the power for operation of the magnets, quite a small motor is used. When the power source is a-c, the motor operates at 3,600 rpm for 60 cycles per second and 12,000 rpm for 400 cycles per second. These motors are synchronous with the line frequency and no speed adjustment is possible. When the power source is d-c, a 10,000-rpm motor is used and an adjustment governor is used for speed adjustment. A 101.69-cycle tuning fork is used for speed-adjustment purposes. The motor speed is the same for all operational speeds, a single gear being changed for the various sending and receiving speeds. (The 25-watt figure given was observed with the d-c power source and the a-c figure may be somewhat higher.)

Printing Out

PAPER-TYPE RELATION

The type cylinder prints from behind the paper and the print hammer presses the inking ribbon against the front of the paper. This unusual arrangement maintains the type face clean at all times and also has the advantage of reducing mechanical wear on the inking ribbon which is struck by the smooth, regular hammer face rather than by a sharp, irregular type face. The main advantage is that the operator has greater visibility of the printing as it is being received. A standard typewriter inking ribbon is used. The ribbon tilts back and downwards after each operation of the print hammer so that the last letter printed can be read. The placement of the type does not affect the ability of the printer to make carbon copies.

CARRIAGE RETURN

Unlike an ordinary typewriter, the paper does not move laterally but only the type cylinder moves. However, the term "carriage return" is still used. Since the type cylinder is quite light (about 1/2 ounce), fast return action is possible. This is important to ensure that the first character at the beginning of the new line is to be clear when only one character space is allowed for the carriage return operation. The MITE is fitted with automatic carriage return and line feed which are triggered when the end of a line is reached without receipt of carriage return and line feed signals. The MITE prints 72 characters per line on standard $8^{1}/_{2}$-inch rolls.

Nonprinting Functions and Special Outputs

The nonprinting functions of the machine include all standard communication printer functions such as carriage return, line feed, shift, signal bell, repeat key, and the break key. Four of these functions can be directly controlled mechanically from front panel buttons. These buttons will not put a signal on the line but can cause the following operations: carriage return, line feed, figures, and letters. In addition, auxiliary outputs can be provided for special functions.

In series with the two pulley-controlled cables which position the type cylinder rotatively and laterally are two slotted plates which slide laterally parallel to one another. The arrangement of the slots in these plates is such that the selection of a nonprinting function will align a certain pair of slots in the two plates. A series of function clutch release levers attempts to sense the alignment of any two slots during each character cycle. When the slots are not in alignment, the levers are deflected. When a nonprinting function signal is received, the appropriate lever is not deflected, but passes through the aligned pair of slots releasing its clutch. These slotted plates can easily be adapted with additional slots and function clutch release levers and clutches, so that additional nonprinting functions can be provided for. There is a position on the type cylinder for each of these functions. Normally, in these positions there is no character on the type cylinder and the print and character advance functions are suppressed. A character could, however, be printed simultaneously with the introduction of a special function if so desired.

In addition, there are about 30 points in the MITE at which contactors, mechanically actuated by the machine, can be installed. These contactors can be used to set up electric circuits in combinations determined by the characters or groups of characters selected by the MITE.

Design and Construction

Design

Certain basic design principles were adhered to in designing and developing this equipment. While miniaturization of the over-all machine was a basic goal, the utilization of full-size components was considered a necessity. The miniaturization of the machine is, therefore, a result of its simplified design and of new approaches to certain problems which permitted that simplification.

Reduction of mechanical speeds within the equipment was considered paramount as was reduction of distances travelled by moving parts. Rotating components were considered preferable to reciprocating components and the weight of moving parts has been kept at a minimum.

The MITE has been designed so that gravity restoration is avoided, permitting operation of the machine in any attitude including upside down. While the standard carrying case accommodates the printer and keyboard, a standard full-size roll of paper, spare parts, and field service tools, other special cases are being developed. The standard roll is $8^{1}/_{2}$ inches wide and $4^{1}/_{2}$ inches in diameter. The machine lends itself quite readily to rack mounting either while still in the carrying case or after removal from it.

Unitized construction was adopted to permit complete replacement of whole sections. Fig. 5 shows the machine broken down into its main subassemblies. These are, from top to bottom, the ribbon magazine, the printer section, main shaft, and the selector section. The breakdown into subassemblies is accomplished with the removal of eight screws and requires less than 10 minutes. The reassembly requires less than 20 minutes. All parts are readily accessible for inspection, lubrication, and servicing. It is estimated that personnel with previous experience in servicing teleprinter equipment can be taught to serive this machine in about 1 week while personnel experienced with typewriters and other business machines can be taught in 2 weeks. There are a total of approximately 70 mechanical adjustments on the machine, many of which are one-time factory adjustments to take up tolerances and which do not require readjustment in the field.

In Fig. 6 the printer, keyboard, and paper roll are shown mounted on a chassis.

Materials

The selection of materials used in construction is in keeping largely with military requirements. Special attention was devoted to those parts where there was a possibility of wear and carboloy inserts were used where necessary. The main structural members are stamped aluminum. Precision castings were used where difficult machining operations could be reduced. Considerable effort was required to determine the most suitable material for the cable in the cable and pulley mechanism.

Many types of chain and metallic cable were tried before it was found that the best material was synthetic fiber cable. This was checked in a test jig for 3,000,000 operations using a test which caused stainless-steel cable to fail after 3,600 operations. After the 3,000,000 operations, there was only minimum evidence of wear.

Companion Devices

Other apparatus under development as companions to the MITE include a miniaturized transmitter-distributor and a reperforator. A stunt box which delivers electric outputs upon certain signals or combinations of signals will also be built as a separate item, independent of the adaptations referred to in the foregoing.

The transmitter-distributor which automatically transmits signals when fed with standard 5-hole perforated tape, will weigh less than 2 pounds including the motor. If the unit is to be used in conjunction with the printer, a common motor will be used and the printer motor will have adequate reserve power to drive the tape transmitter. This additional section will add about 1 inch to the width of the printer and 7 ounces to the weight.

References

1. Trends in Computer Input/Output Devices, J. M. Carroll. *Electronics*, New York, N. Y., Sept 1956.

2. Reference Data for Radio Engineers (book). Third edition, p. 508.

3. A New Look at Communications in the Field Army, H. P. Hutchinson. *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, New York, N. Y., Mar. 1957.

4. Air Force Communication Problems and the Future Air Forces Operational Communication System, C. K. Chappuis. *Ibid.*

5. Digital Printer Boosts Readout Time H. W. Gettings. *Electronics*, June 1957.

Function Tables in Digital Control Computers

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: Function tables of limited size in connection with appropriate interpolation techniques offer considerable advantages in digital computers used in the control loop of servo systems. Several pertinent applications are treated in detail.

RECENT ADVANCES in the art of data storage devices indicate that compact and inexpensive diode matrices or other random access memories will be available in the near future. This revision of present computing and programming techniques is based on this assumption, with the emphasis on stored digital functions.

The concept of stored function tables is not new but, at the present, it is limited to large computer installations with relatively slow access. In this investigation, digital computers in servo loops of weapon systems are treated, primarily considering the severe limitations in computing time and space.

Without dealing with the details of the function table from a hardware point of view, the general feature of the matrix or random access memory under consideration must be that the argument is the address into the decoder matrix, and the stored function can be obtained directly from the encoder without any searching.

Bilog Function

Straight binary notation is predominant in control computers because of the simplicity of the arithmetic and the economy in data storage. There is only one major disadvantage; the variation of significant bits with the magnitude of the expressed quantity. In several cases, this feature is pertinent as in proximity approach, where the quantities vary over several powers of two. In order to maintain a minimum accuracy, a much longer word has to be provided containing a considerable number of ZERO's preceding and following the significant bits. The conventional remedy is scaling or floating radix point, but this technique causes considerable difficulties in control computers, particularly in logic and programming without much gain in computing speed.

A radical solution in this direction can be achieved by using suitable logarithms because of the great savings in computing time and programming for problems of greater complexity. In this case it is possible to justify the scaling operation, if the characteristic of the log is obtained simultaneously.

A floating binary point implies the relation

$$N=2^{s}(1+a_1(2^{-1})+\ldots\ldots+a_n(2^{-n})) \qquad (1)$$

According to this definition, which deviates from the normalized form (e.g. as used in a Remington *1103A*), the quantity N is represented as

$$N=2^{s}F \qquad 1\leq F\leq 2 \qquad (1A)$$

where s is called the characteristic and F is called the fraction. The bounds 1.000 and 2.000 for F are proposed in order to deal with positive mantissas only, in converting N into its log.

$$\log_2 N=s+\log_2 F \qquad (1B)$$

$$M=\log_2 F \qquad (1C)$$

This logarithm of base two expressed in binary notation is the bilog. To compare this concept with straight binary routines depends primarily on the size of the function table.

Taylor's formula with the Lagrangian form of the remainder applied for a bilog function becomes:

$$\log_2(X+x)=\log_2 X+1.44\,x/X-0.72\,x^2/X^2 \qquad (2)$$

The last term must be less than the least significant bit 2^{-n}:

$$0.72\,x^2/X^2\leq 2^{-n}$$

For $X=1.00$ the error would be the maximum and accordingly

$$x\leq 1.18(2^{-n/2}) \qquad (3)$$

This equation is certainly satisfied again for a difference $x\leq 2^{-n/2}$.

In a computer with 16 bit words, the first bit is reserved for the sign and at least three more for the characteristic s. Therefore not more than 12 bits are available for the fraction F. For $n=12$, the increments $x=2^{-6}$ would be small enough for the accuracy specified. In this case, function tables with $2^6=64$ entries suffice.

For the value of X causing the largest error, the numerical interpolation is shown in Table I, wherein octal notation is used instead of binary and denoted by (o). This largest error, shown in Table I, satisfies the condition to be less than the least significant bit 2^{-12}.

From Table II it is seen that the gain in computing time with bilog code increases clearly for higher complexity of the arithmetic operations, particularly if antilog operations are not necessary between individual operations.

Scaling

To scale a binary word into floating notation requires a shift operation to position the most significant ONE with respect to the radix point proper. Instead of shifting the word, the ZERO's preceding the most significant ONE may be sensed and read through a matrix to form the characteristic s. The following relation governs this operation, irrespective of the accuracy:

b = total number of bits in the word $=2^{c}-1$
s' = number of ZERO's preceding the most significant ONE
p = number of bits right of the binary point
s = characteristic as defined in equation 1 requiring c bits

$$s=(b-p-1)-s' \qquad (4)$$

The term in brackets is a constant for the particular digital system. This concept is best illustrated for a typical system, wherein N is represented in several notations as marked.

(d) = decimal
(o) = octal
(b) = binary

$$N=41.75(d)=51.60(o)=101\ 001.110\ 000(b) \qquad (5)$$

Table I. Numerical Interpolation of $\log_2$ 1.0154

	a(o)	b = a	c = log₂b
X	1.0000(o)	1.0000	0.0000
X	0.0077(o)	0.0154	0.0222 = 1.44 x/X
X+x	1.0077(o)	1.0154	0.0220 = true value
	error $2.10^{-4}=0.000064$(o) ≤ 0.0001(o)		

Paper 58-231, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted September 5, 1957; made available for printing February 10, 1958.

E. J. SCHUBERT is with the Burroughs Research Center, Paoli, Pa.

In the straight binary notation, 16 bit words are used; the first bit containing the sign, the binary point in the center of of the word.

$b=15;\ p=8;\ 2^c-1=15\ldots c=4$ (5A)

Hence four bits have to be reserved for the characteristic s with 11 bits remaining to convey the fraction F.

$N=0\ \underline{1}01\ 001.110\ 000\ 00$

$s'=1;\ s=(15-8-1)-s'=6-1=5$ (5B)

$N=2^s(1.30468(d))$ (5C)

As defined in equation 1(A) the fraction F is always greater than 1.00, therefore, there must be always a ONE preceding the radix point. This most significant ONE marked as $\underline{1}$, is therefore a redundant information and may be deleted to increase the number of significant bits to twelve. To discriminate this notation from others, the term "compressed data" is used together with a # to separate the characteristic from the abridged fraction F-1.

$N=5/2340(o)$
$=101\#010\ 011\ 100\ 00(b)$ (5D)

In addressing the bilog matrix, F-1 is required only, and will be substituted by the mantissa M. The characteristic s, which is retained in the word $\log_2 N$, is left from the suppressed ONE #.

Sine-Cosine Tables

A table with a limited number of entries comprises the functions sin X and cos X with full accuracy. The content of the function table are the main values for the range $0 \leq X \leq \pi/4$. For arguments different from the main value X, the functions are interpolated.

In order to determine the number of entries required to interpolate the functions with N significant bits, the Lagrangian remainder in Taylor's formula must be less than the least significant bit.

$\sin(X+x)=\sin X+x(\cos X)-0.5x^2(\sin X)$ (6)

For the value $X=\pi/4$ the last term becomes a maximum.

Table II. Comparison of Binary and Bilog 16-Bit Arithmetic

	Binary	Bilog	Remarks
Sign bit	16th	16th	identical
Range of integer	2^0 to 2^7-1*	2^0 to 2^7 (3 bits)	* binary point between bit 8 and 9
Accuracy	varying 8 to 15 bits	constant 12 bits	
Add/Subtract	1 minor cycle	2 minor cycles	
Multiply	2 minor cycles** 15 minor cycles	6 minor cycles	** whiffle tree
Divide	15 minor cycles	6 minor cycles	
Square root	40 minor cycles	4 minor cycles	

Table III. Numerical Interpolation for Sin 45 Degrees From the Main Value for 44.2 Degrees

	a(o)	b=a	c=πb/4, Radians	d=45.b, Degrees	sin	cos
X	0.7700(o)	0.98438		44.32	0.6986	0.7160
X	0.0076(o)	0.01513	0.01190	0.68	0.0086 = x(cosX)	
φ=X+x	0.7776(o)			45.0	0.7072	
	true value				0.7071	
error 0.0001=0.000032(o)<0.0001(o)						

$\sin \pi/4=0.707$

$0.5x^2(0.707)=0.35x^2 \leq 2^{-n}$ (7)

$x \leq 1.68(2^{-n/2})$

The last equation is certainly satisfied for the difference

$x \leq 2^{-n/2}$ (8)

Accordingly the function table must have $2^{n/2}$ entries. For an argument with 14 bits corresponding to the full angle, the first 3 bits denote the octant only leaving 11 bits for the information $X+x$.

If the argument within the octant comprises 11 bits, then $n=11$, and x must be limited to the least 5 bits. The remaining six bits of the argument belong to the mainvalue of X, to be obtained from a function table with $2^6=64$ entries.

This surprisingly low number of words in the permanent storage suffices as demonstrated for the largest error as shown in Table III. As a matter of convenience, octal notation is used instead of straight binary and denoted by (o) in column a.

Referring to the example mentioned in the preceding with an argument of 14 bits, the most significant 3 bits select the sign of the outputs and control the access to the twin matrix according to Fig. 1 and the following program for $\sin\phi=\sin(X+x)$,

1. Transfer ϕ into address register A with x overflowing into arithmetic unit M.
2. Read cos X from twin matrix T into arithmetic unit M via B.
3. If applicable form lower product x(cos X) and discard.
4. Form higher product x(cos X) and add to sin X read out from T via B.

The program is based on a serial machine with a shift register A and B suitable for serial and parallel in- and outputs. Furthermore, the A-register may complement the content X if so instructed from the control matrix C, according to the octant represented in the three most significant bits of A.

Arc Tan Table

Another application for a rather complex function is the computation of arc tan Z in real time following the cartesian co-ordinates of a flying spot. Two methods are applicable depending on the bounds for x, y. The first method applies for the whole range within the upper and lower bounds $\pm 1/2$ and requires the term $Z=Y/X$, for a spot moving in the direction Y and advancing at the end of this scan in the X direction for one increment only. Within a single line, the X co-ordinate is constant and therefore, the change of Z is limited to the partial differential with respect to Y.

$z=dZ=dY/X \quad \text{for } dX=0$

$Z+z=(Y+y)/X$

Applying Taylor's formula for the single variable Z yields:

$\arctan(Z+z)=\arctan Z+z/(1+Z^2)-z^2Z/(1+Z^2)^2$

The last term corresponds to Lagrangian's remainder with the maximum error at $Z=1.0$ to be less than the least significant bit 2^{-n}.

$z^2(1+z)/(1+(1+z)^2)^2=z^2/2^2(1+z) \leq 2^{-n}$

$z \leq 2^{1-n/2}$

For a range of $\pm 2^{-1}$ for X, Y with an accuracy of 10 bits, the arc tan might be required with 14 bits. The computation of arc tan is performed within one octant only, and the octant itself is denoted by the signs of X, Y, and their relative magnitude. Using three bits for the octant leaves 11 significant bits, correspondingly $n=11$.

$z \leq 2^{-5.5}$ is satisfied for $z=2^{-6}$

hence the range of Z between 0 and 1.00 must be available in 64 main values or in steps of 2^{-6} to achieve the required accuracy.

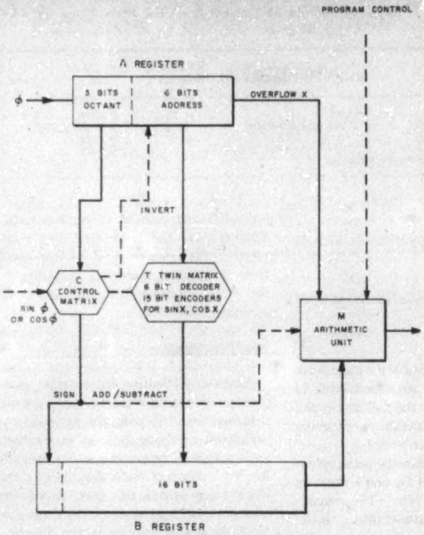

PROGRAM CONTROL
A REGISTER
3 BITS OCTANT
6 BITS ADDRESS
OVERFLOW X
φ
INVERT
C CONTROL MATRIX
SIN φ OR COS φ
T TWIN MATRIX 6 BIT DECODER 15 BIT ENCODERS FOR SIN X, COS X
M ARITHMETIC UNIT
SIGN
ADD / SUBTRACT
16 BITS
B REGISTER

Fig. 1. Sin cos function generator

Data ⟶
Control - - →

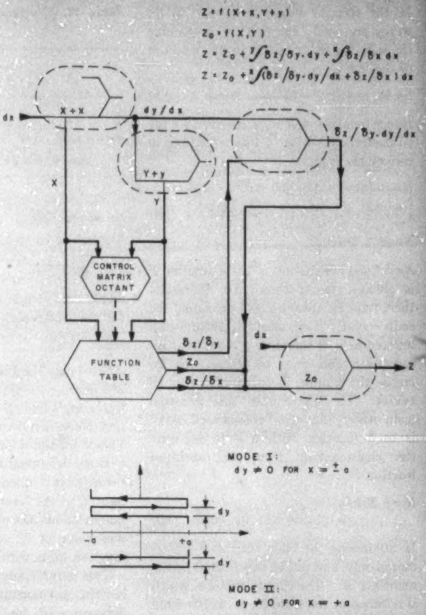

Z = f (X+x, Y+y)
Z₀ = f (X, Y)
Z = Z₀ + ʸ∫δz/δy·dy + ˣ∫δz/δx dx
Z = Z₀ + ˣ∫(δz/δy·dy/dx + δz/δx) dx
dx
X+x
dy/dx
δz/δy·dy/dx
Y+y
X
Y
CONTROL MATRIX OCTANT
FUNCTION TABLE
δz/δy
Z₀
δz/δx
dx
Z₀
Z
MODE I:
dy ≠ 0 FOR x = ± a
dy
-a
+a
dy
MODE II:
dy ≠ 0 FOR x = + a

Fig. 2. Function of two variables

Interpolation is similar to the previous table using an auxiliary matrix for the term $(1+Z^2)^{-1}$ addressed from the same decoder as the main table.

The second method eliminates the periodic division of Y by X in a modification of the Taylor's series for two variables. There is one serious limitation of this more direct approach, the error in vicinity of the origin is excessive. But there are several applications, where data processing might be excluded from this range.

Functions of Two Variables

Instantaneous co-ordinates of the spot are given in x, y, within the bounds $\pm 1/2$. The spot moves from the lower bound in x, to the upper bound in increments of the least significant bit to write a single line. After completing one line, the y coordinate changes for one increment to the adjacent line.

Using Taylor's series again for the appropriate approximation yields:

$$\text{arc tan } (Y+y)/(X+x) = \text{arc tan } Y/X + (Y(x)-X(y))/(X^2+Y^2)+R \quad (9)$$

The last term R is the Lagrangian remainder and determines the number of entries for the mainvalues of X, Y. Instead of calculating X/Y, a matrix with double entries in X and Y is used as decoder and control. Considering the first octant, the flow of pulses corresponding to steps in x increments is accumulated in a binary counter to store $X+x$. The capacity of this counter is 1; accordingly an overflow signal occurs at the upper bounds $+1/2$ and serves as incremental input in the $Y+y$ counter. Three function tables are suggested to reduce computing time according to the following equation:

$$\text{arc tan } (Y+y)/(X+x) = \text{arc tan } Y/X + x(Y)/(X^2+Y^2) - y(X)/(X^2+Y^2) \quad (9A)$$

If speed is a premium, digital integrators are most convenient in connection with the stored function. The first term on the right is the initial condition z_0 of the integrator x, the second term governs the dz/dy integrator, and the third term corresponds to the dz/dx integrator. Counters for X and Y may be substituted by similar integrators controlling the function table. In the conventional symbols for differential analyzers, the block diagram is depicted in Fig. 2. Use of a common high-speed arithmetic programmed to handle data in a random access memory may be applied in larger systems.

In this example, the registers perform linear integration within the bounds set by two adjacent mainvalues as stored in the function table. Initial condition z_0 is extracted from the function table and the two differences dz/dx and dz/dy are added. As soon as the next mainvalue in X or Y has been reached, the content of the z register is discarded and substituted by the accurate initial condition for the new mainvalue. Therefore, accumulation of errors in the z register is inhibited.

Design Equations

To determine the specification of the function tables for the design of a system the following approach is suggested.

Any function $f(X+x)$ may be generated in the manner outlined before providing the argument $X+x$ forms the address into a decoder in such a way that the m bits corresponding to X only are retained in the address portion of the register A, Fig. 1. The function table in the twin matrix T comprises a decoder selecting one-out-of-2^m words contained in the encoder section. This concept is based on uniform increments corresponding to the lowest bit of the address X and nondestructive storage facilities for $f(X)$ and $f'(X)$:

$$f(X+x) = f(X) + x(f'(X)) + x^2(f''(X)/2) \quad (9)$$

With proper scaling the argument X may be restricted to the bounds $+1$ and -1 comprising n significant bits. Linear interpolation will be satisfactory providing the last term in equation 9 is less than the least significant bit 2^{-n}. This condition is certainly satisfied for

$$x^2(f''(X)) = 2^{-n} \quad (10)$$

$$\log_2 x = -n/2 - 1/2(\log_2 f''(X) \quad (11)$$

$$2^m = x(2^n) \quad (12)$$

Hence the number of 2^m words required in the function table is

$$m = n + \log_2 x = (n - \log_2 f''(X))/2 \quad (13)$$

$f''(X)$ should be an extreme within the bounds. The value of $f''(X) = 0$ is irrelevant and indicates straight line intervals. Any symmetry may be utilized in reducing the size of the permanent storage by appropriate control as shown by means of the control matrix C in Fig. 1 depending on intervals of the argument in the A register.

In the subroutine for the interpolation the difference x from the mainvalue X is of concern and might require up to d bits.

$$d = n - m = (n + \log_2 f''(X))/2 \quad (14)$$

Based on the derived expressions the size of the storage facilities can be specified as follows:

1. Address into the decoder section: m bits; equation 13.
2. Storage capacity for mainvalues $f(X)$ disregarding scale factor and sign: 2^m words with n significant bits each.
3. Storage capacity for the first derivative $f'(X)$ if applicable 2^m words with d significant bits each; equation 14.
4. Control matrix for switching certain intervals of the argument depending on the symmetry of the function within the chosen bounds.

References

1. PROGRAMMED VARIABLE RATE COUNTER FOR GENERATING THE SINE FUNCTION, J. N. Harris. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-5, 1956, p. 21.

2. APPROXIMATIONS FOR DIGITAL COMPUTERS (book), C. Hastings, Jr., J. T. Wayward, J. P. Wong, Jr. Princeton University Press, Princeton, N. J., 1955.

3. APPROXIMATING FUNCTIONS FOR DIGITAL COMPUTERS, L. R. Langdon. "Industrial Mathematics," Wayne University Press, Detroit, Mich., vol. 6, 1955, pp. 79–100.

4. DIGITAL COMPUTERS, CIRCUIT AND SYSTEM DESIGN (book), E. J. Schubert. Westinghouse Electric Corporation, Baltimore, Md., 1956, sect. 12.

Some Measurements of Abnormal Corona

GAYLORD W. PENNEY
FELLOW AIEE

JACK G. HEWITT, JR.
ASSOCIATE MEMBER AIEE

CORONA is universally used for charging particles in electrostatic precipitation. In normal corona there is a glow at the active electrode, usually a wire, which supplies ions of one sign to the main charging region. Under some abnormal conditions a glow may exist at points away from the active electrode so that there are ions of both signs in the charging region. These abnormal conditions usually increase the current at a given voltage, and may be quite detrimental in the charging of dust particles. Two types of abnormal corona have been investigated.

The first type of abnormal corona investigated may occur when the active electrode is positive. Insulating material on a positive active electrode may result in a glow which extends out a considerable distance, particularly if the active electrode is large (e.g., 1/8 inch to 3/16 inch) in diameter. This type of abnormal corona will be called a "flare" because of the appearance of the glow. The types of flares discussed in this paper are not intended to represent any particular condition in a normal precipitator. However, it is possible that the same mechanism occurring on the passive positive electrode may be responsible for initiating flashover under many operating conditions. This part of the study was, however, intended only to give a better basic understanding of corona and is not related to any commercial application.

The second type of abnormal corona is due to high-resistivity dust on the large or passive electrode. This produces a glow at the large electrode and results in ions of both signs throughout the charging region and may greatly reduce the charge acquired by particles.

Apparatus and Experimental Procedure

Since the purpose of corona is to charge dust particles, the charge acquired by particles is an excellent means for evaluating corona. For this reason the charge acquired by metal spheres was one of the main tests used. Curves of current versus voltage are another important means for evaluation. Where ionization of air is taking place, visible light is usually emitted. This may be of such low intensity that it is difficult to observe by eye but photographs of corona taken in a darkened room can reveal the location of any glow. Such photographs were therefore used in this study. In order to charge particles effectively corona must have ions of only one sign, the same sign as the wire or active electrode. Ions of a sign opposite to the polarity of the active electrode are very detrimental to the charging process since such ions tend to neutralize the charge on particles. A test for ions of the wrong sign is therefore important in studying any abnormal corona. This test for ions of the wrong sign was made by charging a steel ball with a charge higher than could be acquired from the corona being studied. This highly charged steel ball is then dropped through the corona being studied.

Paper 57-87, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 15, 1956; made available for printing November 13, 1957.

GAYLORD W. PENNEY is with the Carnegie Institute of Technology, Pittsburgh, Pa., and JACK G. HEWITT, JR., is with the International Business Machines Corporation, Yorktown Heights, N. Y.

This investigation was supported in part by the Aerotec and Thermix Corporations and in part by the National Science Foundation; the work was done in part by J. G. Hewitt, Jr., in partial fulfillment of the requirements for the degree of doctor of philosophy at Carnegie Institute of Technology. The authors wish to acknowledge the suggestions and encouragement of E. R. H. Pegg of the Aerotec Corporation and Sidney Orem, Jr., of the Thermix Corporation, and also the work of Samuel B. Craig and David B. Eardley in conducting some of the tests.

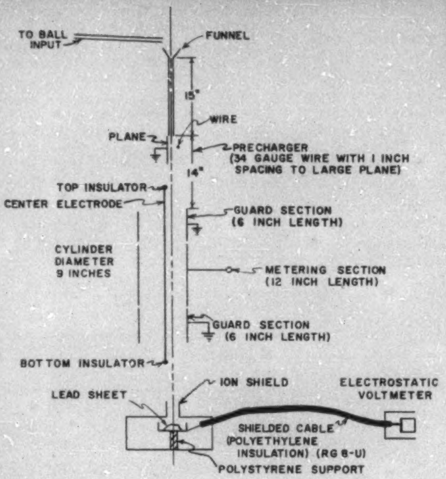

TO BALL INPUT
FUNNEL
15"
WIRE
PLANE
PRECHARGER (34 GAUGE WIRE WITH 1 INCH SPACING TO LARGE PLANE)
14"
TOP INSULATOR
CENTER ELECTRODE
GUARD SECTION (6 INCH LENGTH)
CYLINDER DIAMETER 9 INCHES
METERING SECTION (12 INCH LENGTH)
GUARD SECTION (6 INCH LENGTH)
BOTTOM INSULATOR
ION SHIELD
ELECTROSTATIC VOLTMETER
LEAD SHEET
SHIELDED CABLE (POLYETHYLENE INSULATION) (RG 8-U)
POLYSTYRENE SUPPORT

Fig. 1 (left). Diagram of apparatus for measuring charge and detecting ions of the wrong sign

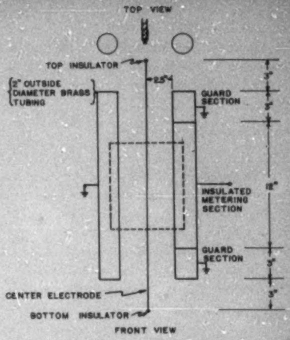

TOP VIEW
TOP INSULATOR
2" OUTSIDE DIAMETER BRASS TUBING
2.5"
GUARD SECTION
3"
3"
INSULATED METERING SECTION
12"
GUARD SECTION
3"
3"
CENTER ELECTRODE
BOTTOM INSULATOR
FRONT VIEW

Fig. 2 (right). Arrangement of electrodes for photographs

If there are no ions of the wrong sign, no loss of charge will occur; conversely, if there are ions of the wrong sign, the charge on the steel ball will decrease markedly.

The principal parts of the apparatus are shown diagrammatically in Fig. 1. The small or active electrode for producing corona consists of a wire or rod mounted along the axis of a 9-inch-inside-diameter cylinder. This 9-inch cylinder serves as the large or passive electrode and consists of a central section 1 foot long connected to a ground through a meter and two end sections which are grounded and serve as guard sections. The central wire or active electrode is supported by insulators and connected to an adjustable-voltage d-c source capable of supplying up to 70,000 volts.

To measure the ability of corona to charge particles, steel balls can be dropped through the corona and caught in a metal cup connected to an electrostatic voltmeter. A grounded metal shield is mounted over the cup to shield the cup from any ions which would otherwise impart a stray charge to the cup. Since the time of fall is relatively long, the charge acquired by the steel balls is substantially the maximum or equilibrium charge. When one knows the equilibrium charge, the field strength can be computed.

This method is somewhat similar to the method used by Cooperman[1] who shot balls at two different velocities and used the resulting data to compute both field strength and current density. In the tests reported here the balls were merely dropped from a height of a few feet which for most cases gave a time of fall which was long as compared to the time to reach the equilibrium charge. This procedure gave more consistent results than did the measurement of the charge acquired by high-velocity balls. However, at high voltages there is some tendency to deflect the ball horizontally; this deflection may be of the order of 0.25 inch at the bottom of the cylinder. But this large deflection would occur only at high field strengths in which case the equilibrium charge is reached very quickly and before any appreciable deflection occurs. Thus, provided the field strength does not increase with radius, the deflection will not produce a serious error if the radius recorded is that at which the ball enters the corona.

The theory for charging spherical particles[2,3,4] indicates that the charge is proportional to the electrical field strength multiplied by the square of the radius of the particles. This theory applies from 1 micron up to a size which materially distorts the electric field. In these tests 3/16-inch-diameter steel balls were used. This size was selected as being large enough so that the balls would drop nearly vertically through the corona region and so that the charge collected would be readily measurable, and yet small enough so that the distortion of the field would not be serious. Near the top of Fig. 1 a "precharger" is shown. This consists of a wire mounted on insulators parallel to a grounded plane electrode. Sufficient voltage can be impressed on this precharger to charge the ball to a higher value than that which could be given by the corona to be studied. This precharge can be measured by dropping the ball with the precharger energized but with no voltage on the wire in the 9-inch

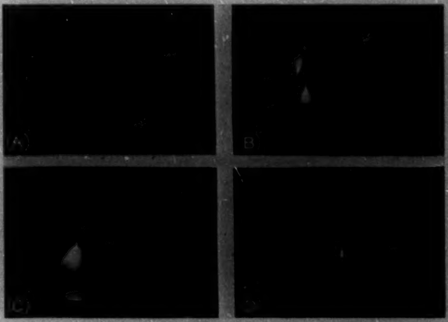


Fig. 3. Photographs of positive corona

A—Clean +48 kv
B—One turn of 3/4 - inch - wide Scotch tape +26 kv
C—One turn of 3/4 - inch - wide Scotch tape +35 kv
D—One turn of 3/4 - inch - wide Scotch tape on wire 0.0063-inch diameter +28 kv
A, B, C—0.125-inch-diameter wire

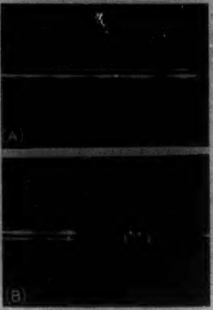


Fig. 4 (left). Photographs of negative corona

A—Clean 0.125-inch-diameter wire at −48 kv, 120 milliamperes per foot

B—0.125-inch wire with one turn of Scotch tape at −48 kv, 135 milliamperes per foot

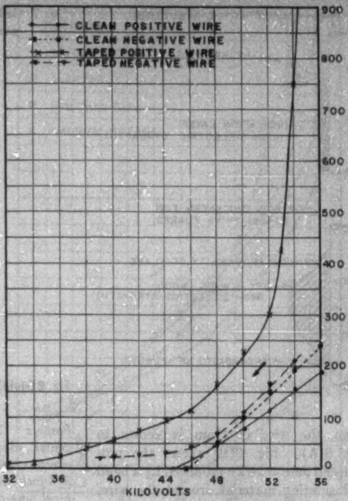


Fig. 5 (right). Volt-ampere characteristics of active electrode abnormal corona

A 3/16-inch-diameter center electrode. Abnormal corona caused by single turn of 3/4-inch-wide Scotch electrical tape on the center electrode

cylinder. If energizing the corona region being studied decreases the charge on the ball, this indicates that ions of the wrong sign are present.

Abnormalities in Positive Corona: Flares

Positive corona as used in the low-voltage 2-stage precipitators is normally a glow along the length of the wire. However, occasionally the characteristic of the discharge may change. A rather faint glow, usually bluish in color, will extend away from the wire. This type of glow will be referred to as a "flare" because in the photographs the glow seems to flare outward from a point on the wire. In cleaning air for ventilating purposes these flares are very objectionable because the ozone generated may be many times that of the normal discharge. Also, wherever there is a glow, one suspects that ions are being formed so that there may be an appreciable number of ions of the wrong sign which would result in poor charging of particles. It seemed desirable to investigate the ability of such flares to charge particles. Furthermore, it is thought that these flares may be responsible for the low breakdown voltage frequently associated with positive corona.

Flares may occur because of some mechanical projection or because of insulating material on the wire. Those starting from insulating material show interesting characteristics; particularly when the active electrode is from 0.1 to 0.2 inch in diameter, the flares originating at the edge of insulating material may give much larger currents than will clean electrodes. It is also easier to obtain reproducible results with these large electrodes than it is with the fine wires usually used in the 2-stage low-voltage precipitators.

This study of flares was carried on to obtain basic information and does not represent any particular operating condition. It is hoped that this and further work may help to explain the low breakdown voltage frequently associated with positive corona, as well as other operating peculiarities.

The visual character of flares can be shown by photographs in a darkened room. To obtain better photographs the high-potential wire was located between two relatively large coplanar cylinders. This arrangement is shown in Fig. 2 with the field of view of the camera shown by the dotted rectangle. Photographs of positive corona are shown in Fig. 3. The uniform corona at 48 kv positive obtained

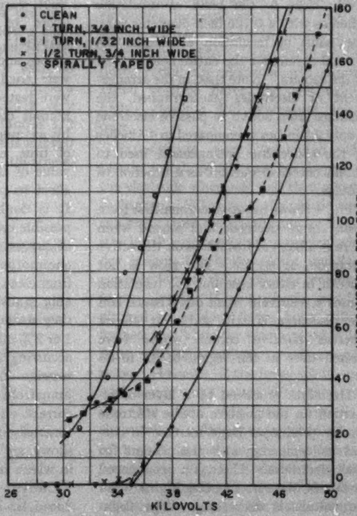


Fig. 6. Volt-ampere characteristics comparing flares caused by different arrangements of insulation

Positive 1/8-inch-diameter center electrode. Scotch electrical tape insulation. Spiral tape winding consisted of 1/4-inch-wide tape wound over the complete electrode length so that 40% of the electrode was covered

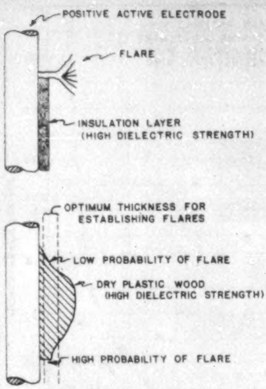

POSITIVE ACTIVE ELECTRODE
FLARE
INSULATION LAYER (HIGH DIELECTRIC STRENGTH)
OPTIMUM THICKNESS FOR ESTABLISHING FLARES
LOW PROBABILITY OF FLARE
DRY PLASTIC WOOD (HIGH DIELECTRIC STRENGTH)
HIGH PROBABILITY OF FLARE

Fig. 7 (left). Conditions for positive flares

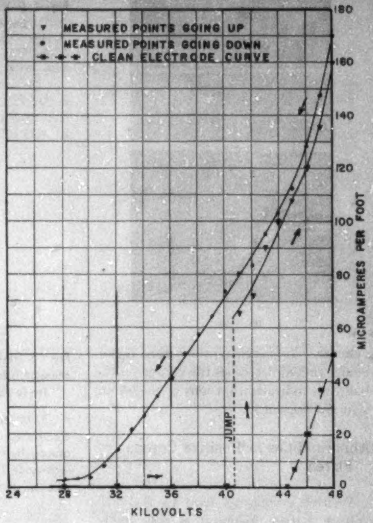


Fig. 8 (right). Hysteresis curve of flare

Positive 3/16-inch-diameter center electrode. Abnormal corona caused by single turn of 3/4-inch-wide Scotch electrical tape on the center electrode

when the wire is clean is shown in Fig. 3(A). Fig. 3(B) shows the pronounced flare at 26 kv produced by a band of insulating material around the wire. In Fig. 3(C) the voltage is raised to 35 kv and the flare has spread across to the opposite electrode. This is very close to flashover. A comparison with Fig. 3(A) for a clean electrode at 48 kv shows the effect of insulating material in reducing flashover voltage in the case of positive corona. Fig. 3(D) shows the much smaller disturbance caused by a band of insulation if the wire is fine. In this case the normal corona resumes relatively close to the band of insulation whereas with a 0.125-inch-diameter wire, the band of insulation eliminates the normal corona for the entire field of view of the camera. Note that, in the arrangement of electrodes photographed, the distance from active to passive electrode is only 2.5 inches as compared to 4.5 inches in the concentric arrangement used to obtain curves of current as a function of voltage.

Fig. 4 shows negative corona first for a clean active electrode and second when there is a band of insulation on the active electrode, at 48 kv. The glow is not smooth in either case but the insulation causes a relatively small disturbance and the discharge is still stable at 48 kv whereas positive corona would have flashed over in the region of 32 to 36 kv.

The flares produced by a layer of insulation on the positive active electrode usually produce a large increase in current as compared with the current for clean electrodes. There is a pronounced hysteresis, the current being higher when the voltage is decreasing. The flashover voltage is relatively low as compared to that for clean electrodes. These characteristics are illustrated in the curves of Figs. 5 to 8.

Volt-ampere curves were taken for a variety of conditions with the intent of gaining an understanding of the nature of active electrode abnormal corona. First, various materials were tried in an effort to discover a dependence on material properties. Second, different shapes and arrangements of one material were tried. Third, different thicknesses were tested. Corona currents under abnormal conditions may vary randomly by as much as 10% over short periods of time such as 1 minute. The mean value of the current was recorded in an effort to present an order of magnitude. It is doubtful, therefore, whether it is possible intelligently to discuss differences of curves which lie within 10 or 20% of one another. The normal corona currents from clean polished electrodes are reasonably stable for short periods of time, as they usually do not vary by more than 1 or 2% during 1 minute. Another effect occurring in volt-ampere curves is hysteresis. The curves have a tendency to jump from essentially zero to a rather high current on the way up and to fall continuously on the way down. Each curve shows an arrow indicating the manner in which it was taken.

Current-voltage curves for flares produced by insulating material on a positive active electrode are very similar for a wide range of materials. Most of the flares were produced by organic insulation since this was easier to handle. To be sure that the phenomena were not due to organic materials, pieces of glass and quartz tubing were placed on the wire and were found to give the same type of current-voltage curve as the organic insulation. With a negative wire a band of insulating material seems to produce only a comparatively small change in the current-voltage curve.

If there are metallic projections from the active electrode, high local gradients will cause corona to start at lower voltages than normal and abnormally high currents will result; but this is true for both polarities. Therefore, the ring of insulating material seems to be necessary to cause the peculiar phenomenon of high positive currents and almost normal negative currents. In order to check the possibility of sharp edges on the insulating material causing high local gradients, which in turn would cause glows and higher currents, two trials were made as follows. First, vacuum grease was applied in a rather thick film (approximately 1/16 inch) to a 3/4-inch length of 3/16-inch-diameter center electrode, and curves similar to those in Fig. 5 resulted. Then the electrode with vacuum grease was held over a dirty gasoline flame so that a complete film of carbon, which is a good conductor as compared to most

VoL
Pt.
195

insulators, was deposited over the vacuum grease. Volt-ampere curves were then retaken. The positive curve was almost exactly the same as the normal curve from clean electrodes, and the negative curve was about normal also. This same procedure was tried with Scotch electrical tape, with similar results. These results indicate that the insulating film causes conditions different from those of conducting projections on the active electrode, so that, when an insulating film is covered with conducting carbon, the abnormal corona ceases and the currents become normal.

Fig. 9. Volt-ampere characteristics comparing flares caused by different insulating materials

Positive 3/16-inch-diameter center electrode. Friction tape, vacuum grease, and Scotch electrical tape applied to a 3/4-inch length of center electrode. Quartz applied to a 6-inch length of center electrode

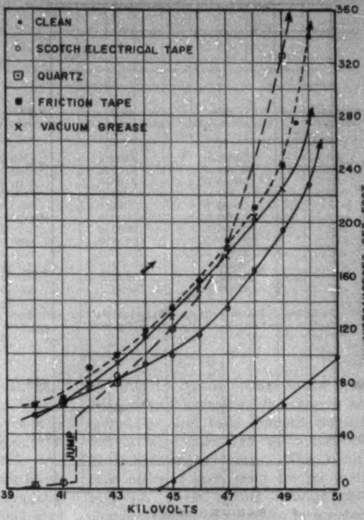


Fig. 6 shows the effect of different sized areas of insulation on the volt-ampere curves of the positive flares. The spirally taped electrode gives the highest currents because flares are established over the complete length of the metering section. The most interesting curves, however, are those in which very small areas of insulation, for example, a 1/32-inch-wide ring of Scotch electrical tape on a 3/16-inch-diameter electrode, causes essentially the same abnormally high currents as do the 3/4-inch-wide rings. The smaller areas of insulation do tend to produce more unstable currents in that there is more variation in current over short periods of time and the curves are not as repeatable as those taken with the 3/4-inch-wide ring of insulation. It is evident from the curves of Fig. 6 that when the area of insulation on the 3/16-inch-diameter active electrode is changed from a 1/32-inch-wide ring to a 3/4-inch-wide ring or to a 3/4-inch-wide half-ring, the voltage–current characteristics remain essentially the same. This fact seems to indicate that the flares are produced at the edges of the layer of insulation. This can be substantiated visually by noting that the stems of the flares always start at the edges of the layer of insulation. Exceptions to this are materials which can break down electrically and form holes; for example, grease and friction tape. The holes can support flares, so that the actual length of edge of the insulation layer is increased by the addition of holes.

The effect of thickness on corona currents from flares is as follows: one layer of tape produces higher currents than do 25 layers. Also, a thicker piece of glass tubing on the center electrode produces less current from flares than does a thinner one. It was found impossible to start flares with thin films of oil or with 50 turns of Scotch electrical tape. Since a zero thickness of insulation and a very large thickness of insulation (about 1/2 inch) will not produce flares, it is implied that there is some optimum thickness. Actually, one layer of Scotch electrical tape about 10 mils thick produces flares more reliably and more readily than do many layers. When grease is applied in varying thicknesses, holes are made due to electrical breakdown, and these holes, which support flares, always seem to occur at thicknesses between 6 to 20 mils. One interesting thing noticed about brass electrodes, however, was that, if they are left for about a month to tarnish in the atmosphere and to pick up dust in a rather dusty room, the electrode will support visual positive flares and will exhibit higher currents and premature sparkover.

Experiments were conducted with plastic wood, which can be shaped on the electrode when wet and then allowed to harden. When wet, the plastic wood can break down electrically and form holes which will support flares; however, when dry, it cannot be broken down, and flares can only form at the edges or at previously formed holes. If wet plastic wood is allowed to produce flares, holes will be formed, and when the plastic wood is dry, flares will be established at these same holes. However, if the holes are tapered with a very small reamer after they were dry, the flares will cease. Fig. 7(A) shows where a flare appears when a layer of high-dielectric-strength insulation is causing it. Fig. 7(B) shows how flares may be initiated and retarded at will by adjusting the shape of the plastic wood on

Table I. Test for Opposite Sign Ions With Flares

A 9-Inch-Diameter Cylinder, 3/16-Inch Center Electrode With 3/4-Inch-Wide Single Turn of Scotch Electrical Tape, 50 Kv

Radius, Inches	Polarity	Precharge, Microcoulombs	Q_N, Microcoulombs	Q_e, Microcoulombs	Opposite Sign Ions
1.4	Positive	0.0015	0.00138	0.00082	Very few
	Negative	0.00146	0.00145	0.00103	No
1.8	Positive	0.00134	0.00130	0.00073	Questionable
	Negative	0.00157	0.00155	0.00058	No
2.2	Positive	0.00150	0.00150	0.00065	No
	Negative	0.00155	0.00153	0.00052	No
2.6	Positive	0.00151	0.00150	0.00068	No
	Negative	0.00160	0.00157	0.00051	No
3.0	Positive	0.00151	0.00149	0.00062	No
	Negative	0.00159	0.00160	0.00042	No

Q_N = The net charge remaining after the precharged ball is dropped through the corona being studied. If the charge Q_N is measurably less than the precharge this is an indication of ions of the wrong sign.
Q_e = The charge imparted to an initially uncharged ball by dropping through the corona being studied.

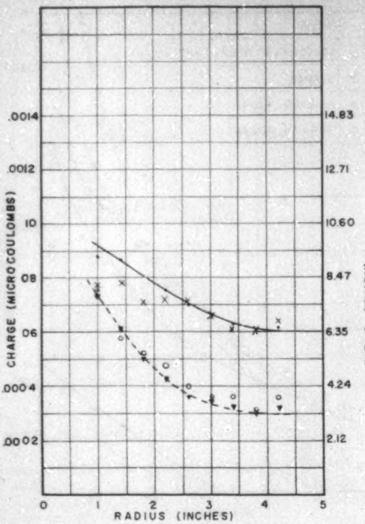


Fig. 10 (left). Charge acquired by 3/16-inch-diameter steel balls from drop through corona discharge

A 3/16-inch-diameter center electrode
▼ Clean electrode (+48 kv). The following have a single turn of 3/4-inch-wide Scotch electrical tape on the center electrode:
●(+48kv) May 23, 1956
X(+48 kv) April 21, 1956
O(−49 kv) June 19, 1956

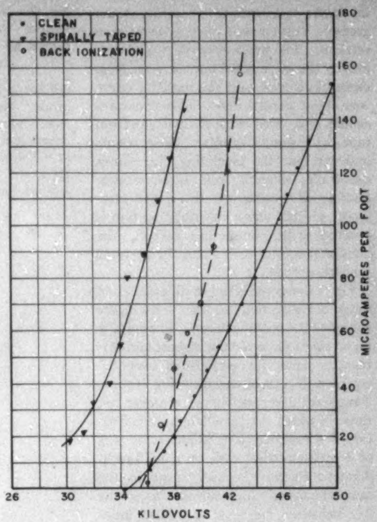


Fig. 11. Volt-ampere characteristics comparing flare and back-ionization

Positive 1/8-inch-diameter center electrode. Flare caused by spiral tape winding consisting of 1/4-inch-wide tape wound over the complete electrode length so that 40% of the electrode is covered. Simulated back-ionization caused by punched cambric layer on the passive electrode

the center electrode. (Assume the plastic wood will not break down electrically, as is generally true when it is dry.) In Fig. 7(B) a flare will not start at the top of the plastic wood, because the surface tapers gradually into the electrode. However, at the bottom the flare will start because of the perpendicular surface near the optimum thickness. It is possible to start and stop flares at will by adjusting the shape of the plastic wood as described; however, it seems that the perpendicular surface supporting the flare has to be quite smooth. If the surface is roughened with a knife or drill, flares are reluctant and sometimes impossible to start. Once a flare has been started, however, it is easy to restart it. It may be possible that the active region of glow about the surface tends to smooth off the jagged edges, thereby making it easier to start the next flare. There can be no such control of flares when the insulating layer breaks down electrically, as with, for example, grease and friction tape.

Fig. 8 shows the hysteresis effect of the voltage–current curves of the positive flare. The hysteresis curves are repeatable if sufficient time is allowed after the current goes to zero. In the tests reported here, about 1 or 2 minutes must be allowed at zero current in order to repeat the curve of Fig. 8 over and over. If the time is not taken, the large current jump occurs at a lower voltage, and the shorter the time the lower the voltage until there may appear to be negligible hysteresis. The arrows on the graphs tell the direction in which the voltage was being varied.

A very important characteristic of volt-ampere curves of abnormal active electrode corona is the tendency for premature sparkover or arcing of positive polarity. For a wide range of sizes of the center electrode in the 9-inch-diameter cylinder, if some insulating layer causes a flare, sparkover occurs between 50 and 55 kv. There seems to be little correlation between breakdown voltage and center electrode size. Different insulating materials seem to cause only very slight changes in breakdown voltage. (Fig. 9 gives volt-ampere characteristics comparing flares caused by different insulating materials.) Negative abnormal corona will not cause the excessively premature breakdown that positive flares will. In fact, the breakdown voltage of negative abnormal corona on the 9-inch cylinder was always above 70 kv which was the capacity of the source used. It is interesting to note from Peek's studies[5] that the order of magnitude of sparkover voltage for clean electrodes for a 9-inch-diameter cylindrical geometry is above 100 kv.

Where there is a glow, one suspects that ions are being formed, and that ions of the unwanted sign may interfere with the charging of particles. In order to study this question, precharged balls were dropped through the corona region using the apparatus shown in Fig. 1. The results of a test at 50 kv are shown in Table I. There was a slight indication of ions of the wrong sign at a radius of 1.4 inches, but at radii over 2 inches no evidence of ions of the wrong sign was found. It is possible that there may have been some point on the circumference with a more active glow and a significant number of ions of the wrong sign. However, if such a region exists it was not found in a number of tests and so would not appear to have any large effect on the average charge imparted to particles.

Measurements of the charge imparted to steel balls show that at a given voltage corona from clean electrodes does not give as high a charge as the corona which accompanies flares. Since for larger radii there are no ions of the wrong sign and since the balls are reaching substantially the steady-state charge, it appears reasonable to plot both the charge and corresponding computed field

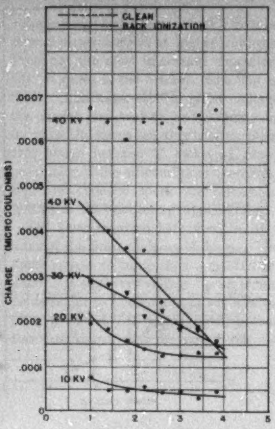


Fig. 12. Charge acquired by 3/16-inch-diameter steel balls from drop through simulated back-ionization

Positive 0.0063-inch-diameter center electrode. Relative humidity 48%

Voltage, Kv	Ratio of Currents, Back-Ionization to Normal
10	0.41
20	0.45
30	1.41
40	1.80

strength. This is done for a series of tests in the curves of Fig. 10. These curves indicate that flares give a marked reduction in gradient near the wire and thus increase the gradient at larger radii. Since 80% of the charging volume lies outside of a 2-inch radius, the effective charging ability is considerably increased for a given voltage. Of course, from a practical viewpoint this gain is usually nullified by the reduced sparkover voltage which accompanies the flares.

Back-Ionization

Back-ionization is a major trouble in many precipitator applications. Photographs showing the glow on the large or passive electrode and curves of current as a function of voltage have been published in a previous paper.[6] The work described here attempts to study the charging of particles and to determine under what conditions ions of the wrong sign exist in this type of abnormal corona.

Table II. Test for Opposite Sign Ions; Simulated Back-Ionization

A 9-Inch-Diameter Cylinder, 0.0063-Inch Center Electrode; Positive Polarity; Relative Humidity 48%

Voltage (Kv) and Current (Microamperes per Foot)	Radius, Inches	Precharge, Microcoulombs	Q_N, Microcoulombs	Q_e, Microcoulombs	Opposite Sign Ion
10	1.8	0.0014	0.0014	0.00006	No
2.3	2.6	0.0014	0.0014	0.00004	No
	3.4	0.0015	0.0015	0.00004	No
20	1.8	0.0014	0.00016	0.00015	Yes
28.5	2.6	0.0014	0.00019	0.00012	Yes
	3.4	0.0015	0.00019	0.00013	Yes
30	1.8	0.0014	0.00023	0.00026	Yes
188	2.6	0.0014	0.00024	0.00022	Yes
	3.4	0.0015	0.00021	0.00018	Yes
40	1.8	0.0014	0.00039	0.00036	Yes
600	2.6	0.0014	0.00035	0.00023	Yes
	3.4	0.0015	0.00033	0.00019	Yes

Q_N = The net charge remaining after the precharged ball is dropped through the corona being studied.
Q_e = The charge imparted to an initially uncharged ball by dropping through the corona being studied.

In back-ionization with actual dust deposits, it is difficult to obtain reproducible tests. Some tests were made in which the 9-inch cylinder was lined with varnished cambric. The cambric was thick enough so that it would not break down under application of the available voltage to the wire. A large number of pin holes were punched through the cambric before it was cemented to the cylinder. This gave a high-resistivity lining which could form a glow at each of the punched holes. There was some question that the cambric might carbonize at the holes and change the characteristic and some blackening at a few holes could be observed, probably where arcover occurred. However, the "simulated" back-ionization obtained using this perforated varnished cambric gave more reproducible curves than when actual dust layers were tested. These tests simulated the condition of a very thick layer of high-resistivity dust on the large or passive electrode.

When the corona current for clean electrodes is compared to the current with high-resistivity dust on the large electrode as in Fig. 11, there is usually a region at low voltage where the current for the clean electrode is higher than for the condition where the large electrode is covered with high-resistivity dust. At some point the curves cross and above this crossover the current with back-ionization increases much more rapidly than the current with clean electrodes. The high-current condition is explained by the fact that ions of the wrong sign tend to neutralize space charge and in this way increase current. It might seem that a current below normal would have no ions of the wrong sign and that this low-current condition would effectively charge particles.

In order to investigate the possibility of an optimum corona voltage in the presence of back-ionization, the charges acquired by steel balls dropped through corona were measured as a function of radius and voltage. The results of a series of such tests are shown in Fig. 12. These results indicate that charging cannot be improved by a reduction in the voltage. Another interesting feature is the effect of radius on the reduction of charge in the presence of back-ionization. The reduction in charge brought about by back-ionization is much greater near the outer electrode than it is near the inner electrode. This indicates that there is a higher percentage of ions of the wrong sign near the outer electrode.

In order to detect ions of the wrong sign, the precharger was energized to impart a high charge. In this test the precharged balls are dropped through the corona being studied and the resulting or final charge measured. If the charge remaining on the balls is equal to the precharge, then it is assumed that there are no ions of the wrong sign in the corona being studied. On the other hand, a marked decrease in charge indicates that there are ions of the wrong sign.

The results of one series of tests is shown in Table II. For this series a small wire, 0.0063 inch in diameter, was used so that at 40 kv and with clean electrodes the variation of voltage gradient with radius was quite small. With clean electrodes at 40 kv the charge was about 0.65×10^{-9} coulombs per ball over the range of radii that could be measured. The precharge used was from 1.4 to 1.5×10^{-9} coulombs so that it was well above any charge that could be obtained from the corona being studied. At 10 kv the current is only 2.3 microamperes and at this low current there is no

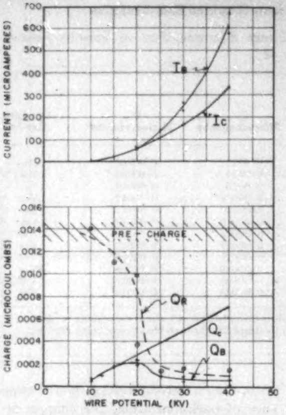


Fig. 13. Comparison of corona from clean electrodes with back-ionization due to a thin layer of dust

I_a=current with clean electrode

I_b=current with thin layer of dust of 2.5×10 inches per ohm-centimeter

Q_R=resultant charge when precharged ball is dropped through corona with back-ionization

Q_c=charge imparted to steel balls by corona from clean electrodes

Q_B=charge imparted to steel balls with back-ionization

evidence of ions of the wrong sign; however, the charge acquired by the ball at this low voltage is so small that it is of little value. At 20 kv the current is 28.5 microamperes as compared to 65 microamperes for clean electrodes or 44% of normal, but ions of the wrong sign are present in sufficient numbers to reduce the precharge almost to the charge that would be obtained without precharge. Of course, at any higher voltage there is greater evidence of ions of the wrong sign.

Tests were next made with a thin layer of dust precipitated onto the inside of the passive electrode. The apparatus was enclosed so that it could be heated slightly above room temperature in order to control relative humidity and resistivity. The dust used was fly ash from a station where considerable trouble had been experienced with precipitators. Resistivity was measured with the apparatus described in a previous paper[6] at the same temperature and humidity as was used for the corona measurement. With this thin layer of dust (less than 1/32 inch thick) the current measured with the high-resistivity dust was never significantly lower than that with clean electrodes. With a dust layer at a resistivity of 2.5×10^{11} ohm-centimeters and at 20 kv the current was normal but there was a definite indication of ions of the wrong sign. As the voltage was increased the corona current tended to be above normal and the loss of charge from a precharged ball was greater, indicating an increase in the percentage of ions of the wrong sign. Results obtained by dropping balls at a radius of 3 inches are plotted in the curves of Fig. 13. The upper curves show the current both for clean electrodes and for a thin layer of high-resistivity dust on the outer electrode. The plot of charge acquired by a 3/16-inch sphere with clean electrodes rises steadily as the applied voltage increases. However, with a thin layer of fly ash having a resistivity of 2×10^{11} ohm-centimeters the charges for 10 kv and 15 kv are nearly the same as for clean electrodes. Between 15 and 20 kv the charge acquired by spheres reaches a maximum and in the higher voltage range decreases to a very small value.

The results shown in Fig. 13 indicate that corona current is not necessarily a reliable indication of the charging ability of corona. For example, at 20 kv the current is normal and yet the charge has dropped off appreciably as compared to that for the clean condition. From the results of the test for ions of the wrong sign, curve Q_R of Fig. 13, it appears that at 10 kv a sphere precharged to 14×10^{-10} coulombs lost over 20% of this precharge in dropping through the corona region. This indicates a significant number of ions of the wrong sign even though the effect on the acquired charge is small. At 30 kv about 90% of the precharge is lost in traversing the corona region and the charge acquired by an initially uncharged sphere is less than 20% of that which would be received if the electrodes were clean.

In the series of tests shown in Fig. 14 the voltage was held constant and the radius at which the balls were dropped was varied. With a fine wire as the active electrode the charge acquired was nearly independent of radius if the resistivity of the dust on the outer electrode was low. However, if the temperature was raised so as to reduce the relative humidity and increase the resistivity, the charge acquired was reduced at all radii but the reduction was much greater near the outer electrode. This agrees with the tests simulating the thick dust layers and indicating marked recombination of ions.

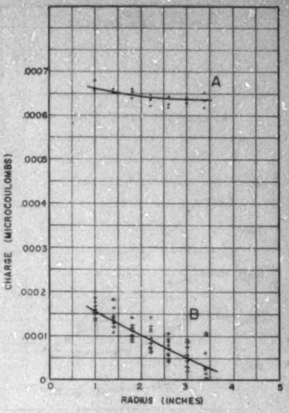


Fig. 14. Charge as a function of radius

0.0063-inch-diameter wire at +40 kv

Thin layer of fly ash on outer electrode

A—High relative humidity

B—Low relative humidity

One outstanding characteristic of the curves of Figs. 12 and 14 is the decrease in charge acquired by a sphere dropped at a large radius as compared to one dropped at a small radius. The curves of Fig. 12 indicate that this characteristic is much more pronounced at high voltages. At a large radius the charge increases very little as the voltage increases. In contrast to this, if the spheres are dropped at a small radius the charge acquired increases much more rapidly as the applied voltage increases. This indicates that at the high-current condition obtained at the higher applied voltage a marked recombination of positive and negative ions is occurring. Near the outer electrode where the positive ions are generated the fraction of positive ions is high and the resulting charge acquired by particles is low. Nearer the wire a large fraction of the positive ions have recombined with negative ions so that there are relatively few ions of the wrong sign and the charge acquired by particles is a relatively large fraction of that which would be acquired if the electrodes were clean.

Discussion of Results

Flares caused by insulating materials on the positive active electrode result in a large increase in current in the region near the voltage at which corona starts. High-resistivity material on the passive electrode with both polarities may show a distinct decrease in current in this low-voltage range but a current which may be several times normal at higher voltages. Probably the most important characteristic of flares is the marked reduction in flashover voltage. At a given voltage the charge imparted to particles may be increased by flares but this is of no practical value because of increased ozone generation in the air-conditioning application and reduced sparkover voltage in industrial dust-collecting applications.

The tests reported here describe only some of the characteristics of flares. The flares appear to cause such a marked reduction in flashover voltage that it would be desirable to know more of the mechanism involved. It seems possible that this mechanism may be responsible for flashover at relatively low voltage in other types of equipment.

The tests with high-resistivity dust indicate that corona current is not a reliable indication of the ability of the corona to charge particles. Even in a short length of one wire the current can be normal and yet have a high percentage of ions of the wrong sign, which results in poor charging. In a larger unit one section of wire may have a high current and another a low current, which produces a normal total current but very poor charging. In a single-stage precipitator the rate of precipitation depends upon the charge on particles near the outer electrode. Tests show this charge can approach zero with bad back-ionization.

References

1. A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation, P. Cooperman. *AIEE Transactions*, vol. 75, pt. I, Mar. 1956, pp. 64–67.

2. Charging of Spherical Particles in an Ionizing Field, M. Pauthenier, M. Moreau-Hanot. *Journal de Physique et le Radium*, Paris, France, series 7, vol. 3, 1932, pp. 590–613.

3. Method of Size Measurement for Suspended Particles, H. Rohmann. *Zeitschrift für Physik*, Berlin-Charlottenburg, Germany, vol. 17, 1923, pp. 253–65.

4. Particle Charging in Electrostatic Precipitation, H. J. White. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1186–91.

5. Dielectric Phenomena in High Voltage Engineering (book), F. W. Peek. McGraw-Hill Book Company, Inc., New York, N. Y., 1929, p. 115.

6. Electrostatic Precipitation of High-Resistivity Dust, G. W. Penney. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1192–96.

High-Efficiency Push-Pull Magnetic Amplifiers with Transistors as Switched Rectifiers

A. G. MILNES
MEMBER AIEE

IN MANY magnetic amplifiers, push-pull action is used to obtain outputs that depend in polarity (or phase) on the signal polarity (or phase).[1,2,3]

If polarity-reversing d-c output is required, series or T-summing circuits may be used, but considerable power is lost in such networks and the theoretical efficiency has a maximum value of 17%. Improvement is possible if the load has twin windings, since then magnetic summing may be adopted. The effective use of the load winding space is halved by the need for two load windings, only one of which is energized in the full output condition, and therefore the efficiency of magnetic summing is not more than 50%. For d-c loads where twin load windings are not practical, a bridge configuration has been described that requires twin windings on each core of the amplifier.[1,4,5] A dummy load is necessary for reset and ballast purposes and this reduces the circuit efficiency to 50% or less. In motor-type applications there are difficulties in matching the dummy load to the machine load under varying speed conditions, and a compromise must be made that involves additional sacrifice of efficiency.[4]

In this paper, various circuits are considered that use transistors as switched rectifiers and efficiencies well in excess of 50% for ± polarity-reversing d-c outputs are obtained. Alternating-current output circuits are not examined, since in general these do not incur the efficiency losses of d-c output circuits. The circuits on which the work is based are the push-pull voltage-reset (Ramey) amplifiers proposed by Maine, in which the circuit elements are coupled by switched rectifiers that prevent circulating currents from bypassing the load.[6] A typical full-wave push-pull circuit is shown in Fig. 1(A) and this has input/output characteristics as in Figs. 1(B) and 1(C). With a d-c signal of one polarity, cores *1* and *4* provide output in sucessive half-cycles while cores *2* and *3* are "passive"; and for the other signal polarity the action interchanges and cores *2* and *3* provide output of reversed polarity. With an a-c signal, cores *1* and *2* provide a-c output and *3* and *4* are passive; and reversal of the signal phase brings about reversal of the core actions and the phase of the a-c output reverses. The switched rectifiers *a*, *b*, *c*, *d* in the load side are controlled by bias voltages of gate-supply magnitude acting through resistors R_z equal to the load resistance. This ensures rectifier cutoff for half of each cycle and limits the rectifier current flow to the load current magnitude. With these necessary proportions the switching power for the load circuit is twice the maximum load power, and hence the maximum circuit efficiency is 33% (assuming the power loss in the control side of the circuit to be small).

The concept that suggests itself for the improvement of efficiency in this circuit is the substitution of transistors for the switched rectifiers since the functions of half-cycle cutoff and current limiting can then be performed for the expenditure of much reduced power.

Push-Pull Reset Circuits with Transistor Switches

In switched operation of a transistor the base-drive current need be only equal to the output current divided by the current gain of the unit, which is normally 10 or more. The base-drive voltage must be sufficient to supply this base current in a typical switching configuration and this is perhaps merely 2 or 3 volts. Thus the base-drive power may be only a few per cent of the power controlled as a switch.[7] The voltage drop between emitter and collector when acting as a closed switch

---

Paper 58-77, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 7, 1957; made available for printing November 7, 1957.

A. G. Milnes is with the Carnegie Institute of Technology, Pittsburgh, Pa.

The work described in this paper was performed under an Office of Naval Research contract *NONR 760(09)*.

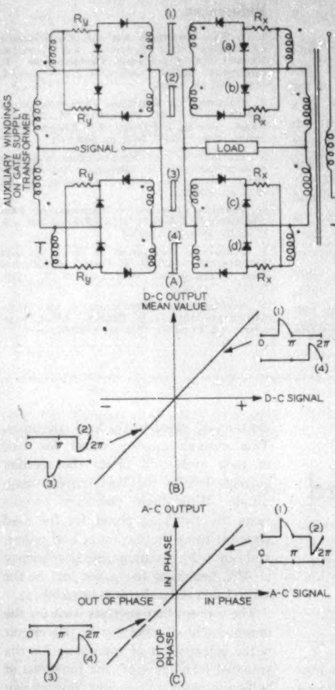


Fig. 1 (left). Push-pull reset magnetic amplifier with switched rectifiers giving d-c/d-c or a-c/a-c characteristics

A—Full-wave circuit (Maine)
B—Action with d-c signal
C—Action with a-c signal

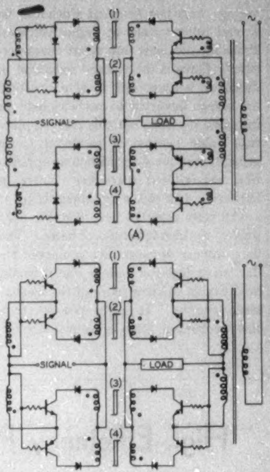


Fig. 2 (right). Push-pull reset magnetic amplifiers with switched transistors

A—Switched rectifiers in load side of Fig. 1 replaced by transistors
B—With transistors in both the control and the load sides

may be 0.3–1.0 volt at full-load current depending on the transistor type and the amount of drive provided. The total switch loss therefore is unlikely to exceed 5% of the power switched in a correctly proportioned circuit.

Rectifiers *a*, *b*, *c*, and *d* of Fig. 1(A) may be replaced by transistor switches as shown in Fig. 2(A). Then for 5% switch loss per pair of transistors the theoretical maximum efficiency of Fig. 2(A) should be about 90% as compared with 33% for the earlier circuit. Other losses that must be allowed for include winding resistance losses and normal rectifier forward-voltage losses, and these may amount to 10 to 20% depending on the design proportions. The power supplied to the control circuit for flux reset is normally small, but also may be allowed for in assessing efficiency.

The circuit of Fig. 2(A) may be rearranged to use *p-n-p* and *n-p-n* units in complementary symmetry, Fig. 2(B), which allows base drive to be obtained from the gate-voltage windings, thus simplifying the supply transformers. In Fig. 2(B) transistor switches are shown also in the control side of the circuit, although the power levels there are low and the effect on efficiency correspondingly small.

The circuit of Fig. 2(A) gives reversing-polarity d-c output for a d-c signal and reversing-phase a-c output for an a-c signal input. Rearrangement of the control circuit as shown in Fig. 3 interchanges the control action so that a d-c signal gives a-c output and vice versa. Although only full-wave circuits are given, the same principles apply to half-wave versions of these circuits.[6]

Since the circuits are constructed from voltage-reset-(Ramey-) type configurations the transient response to be expected is merely the half-cycle delay of such arrangements.

The switch-off condition for a *p-n-p* transistor is that the base must be positive with respect to both of the other electrodes. Thus the magnitude of the switch-off voltage between the emitter and the base limits the voltage that can be sustained with the collector positive with respect to the emitter.[7] This is the reason for the inclusion of diode rectifiers in series with each transistor, and the possibility exists of eliminating the diodes if the base-switching voltage can be made large.

Experimental Examination of the New Push-Pull Reset Circuits

An experimental examination was made to establish that the full-wave circuits of Fig. 2 and Fig. 3 functioned as expected and gave high efficiency. Four cores of Orthonol of $2^1/_4$-inch inner diameter, $3^1/_2$-inch outer diameter, and 3/4-inch tape width were provided with gate windings of 500 turns and control windings of 166 turns. At 60 cycles per second the gate-winding-saturation voltage per core was just over 50 volts (rms). In these push-pull circuits, triggering of an active core produces voltage across the load which adds to the gate-voltage supply of the passive section of the amplifier and therefore the gate-voltage rating in push-pull must be half that for normal single-ended operation. In these experiments the gate voltage of each half of the circuit was therefore set at 25 volts. The power transistors available were *2N68* units, and these were selected to have collector-emitter turnover voltages in excess of 60 volts when the base was positive with respect to the emitter. For the circuit of

Vol.
Pt. 1
1958

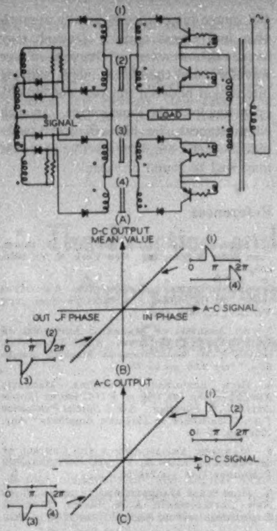


Fig. 3. Push-pull reset amplifier with transistors and with control circuit arranged for d-c/a-c or a-c/d-c characteristics

A—Circuit: compare the control side with Fig. 2(A)
B—Action with a-c signal
C—Action with d-c signal

Fig. 2(B), *n-p-n* transistors are also required and type *2N95* units were used.

Examination of the circuit waveforms confirmed the expected action and the improvement in efficiency. A typical characteristic for the circuit of Fig. 3 for a load resistance of 150 ohms and gate and reset voltages of 25 and 9 volts, respectively, is given in Fig. 4. The circuit efficiency depends on the output level and increases as this approaches full output. At a signal level of 9 volts, corresponding to almost full output in Fig. 4, the efficiency represented by the load-power output divided by the input power was 80%. The need for high-impedance wattmeters to measure true power may be avoided by taking other ratios, such as that of the load current to the total input current, to represent the effectiveness of the push-pull arrangement. For the characteristic of Fig. 4 a load current of 124 milliamperes required an input current to the gate side of 144 milliamperes (mean) and an equivalent of 25 milliamperes to the control circuit side. The ratio of load to input current was therefore 0.76 and the ratio of output–input volt-amperes was 60%. This represents a considerable improvement over values of less than 17% for conventional circuits and the values of less than 33% for switched rectifier circuits.

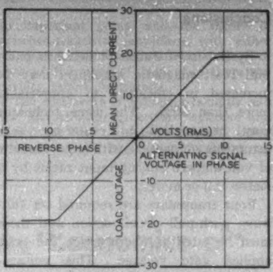


Fig. 4. Typical experimental output characteristic for the circuit arrangement of Fig. 3

For these experiments the amplifier elements were designed to be of negligible winding resistance to facilitate efficiency studies; the number of turns was low and the input impedance was correspondingly low (about 700 ohms). This had the effect of limiting the power gain in the experiments; with a load of 150 ohms the gain was about 20 and for a 50-ohm load the gain was about 60. In any practical application the gain could be improved by redesign of the core elements.

Push-Pull Center-Tap Self-Saturating Circuit of High Efficiency

The voltage-reset (Ramey) amplifiers that have been examined in the high-efficiency push-pull configurations have minimum output for zero signal. Conventional self-saturating amplifier circuits do not have this type of characteristic, although by added bias the output at zero signal may be reduced to a low value. The advantage of this condition is that for any given direction of signal two cores of the push-pull configuration do not contribute to the output and therefore the fluxes do not saturate and thereby open up circuit paths through which circulating current may bypass the load. With this consideration in mind it was decided to examine the transistor technique of push-pull coupling with conventional self-saturating circuits suitably biased to be of low standing current at zero control signal current.

The circuit arrangement examined, Fig. 5(A), is identical on the load side to the configurations (Fig. 2(A) and 3) already discussed. On the control side the bias and signal circuit connections are such that cores *1* and *4* may be regarded as forming a center tap self-saturating circuit acting in push-pull with a similar configuration formed by cores *2* and *3*. In Fig. 5(B) the full-line curve is a typical push-pull characteristic and the dotted line represents the output obtained from cores *1* and *4*, with the other cores disconnected to show the extent of the bias provided. The gate voltage was 25 volts and loads in the range of 150 to 50 ohms were used. For a load current of 290 milliamperes in 50 ohms the total input current to the circuit was 325 milliamperes (mean). This represents a volt-ampere output/input ratio of 0.52, which is very much better than for conventional series or T-summing networks.

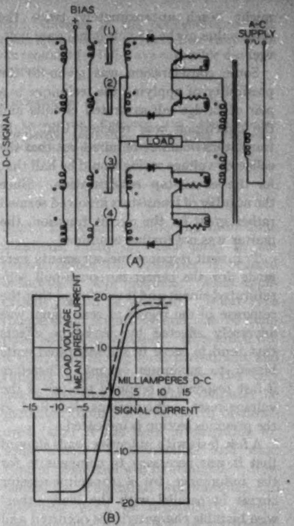


Fig. 5. Push-pull self-saturating amplifier with switched transistors

A—Circuit of push-pull center-tap amplifier
B—Experimental output characteristics

With Orthonol cores, there was a tendency for looping of the characteristic to occur at low signal levels. Triggering is a well-known effect in this class of core material, but to confirm that the circuits were not responsible the experiments were repeated with *4-79* Mo-permalloy cores and the effect was absent. Inspection of the waveforms for both types of material showed that the reverse voltages across rectifiers and transistors might reach approximately twice the peak value of the 25-volt (rms) gate voltage.

Some consideration was given to the possibility of applying the technique to a pair of bridge self-saturated circuits and the conclusion was reached that eight transistors would be required but that the collector voltage rating would be half that for the center-tap configuration. Since the number of transistors involved seemed rather great for the circuit function, the matter was not further examined.

Transient response measurements were made for the center-tap push-pull self-saturated circuit and it was found that the response in the region of zero signal was adversely affected by dead-time effects that tend to occur in self-saturated units biased to minimum output. Therefore if fast response is required the use of the voltage-reset configurations described in the previous section is indicated.

A few tests with inductive loads showed that it was necessary to compensate for the inductance by a capacitor-resistor circuit in parallel with the load, otherwise bistable characteristics occurred and circulating currents could flow between the push-pull units.

Conclusions

By the use of transistor switches, push-pull reset magnetic amplifiers may be increased in efficiency to appreciably more than 50%. The same technique may be applied to center-tap self-saturating connections with improvements of volt-ampere output/input ratios by a factor of 3 or more.

Four transistors are required for full-wave push-pull circuit action and these must be rated at about twice the peak applied gate voltage. This voltage-rating requirement constitutes at present a limitation on the applicability of the technique. The power dissipation in the transistors is small and temperature effects on linear parameters are not important since switched action is used.

In many magnetic amplifier applications the low volt-ampere efficiency of conventional push-pull configurations is not important. In certain military applications, however, the power supply available is limited and excess current demand must be avoided to conserve power and to restrict unnecessary heat generation that may present dissipation problems. With a limited power supply the source impedance is usually appreciable and excess current demand may cause deterioration of supply-voltage waveform and therefore adverse coupling interaction between independent circuits operating from the same source. In such circumstances the high-efficiency push-pull magnetic amplifier circuits proposed may well be found of value.

References

1. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, chap. 26.

2. TRANSDUCTORS AND MAGNETIC AMPLIFIERS (book), A. G. Milnes. St. Martin's Press, New York, N. Y., 1957, chap. 7.

3. AN ANALYSIS OF MAGNETIC AMPLIFIERS IN PUSH-PULL, D. L. Critchlow. *AIEE Special Publication T-86*, "Conference on Magnetic Amplifiers," July 1956, pp. 52–73.

4. HIGH EFFICIENCY PUSH-PULL MAGNETIC AMPLIFIER AND ITS USE AS A D-C SHUNT MOTOR DRIVE, C. L. Boyajian. *AIEE Special Publication T-98*, "Conference on Magnetic Amplifiers," Aug. 1957, pp. 35–44.

5. MAGNETIC AMPLIFIERS FOR THE CONTROL OF SERVOMOTORS, Allen R. Perrins. *AIEE Special Publication T-86*, pp. 246–54.

6. HIGH SPEED MAGNETIC AMPLIFIERS AND SOME NEW DEVELOPMENTS, A. E. Maine. *Electronic Engineering*, London, England, May 1954, p. 180.

7. JUNCTION TRANSISTORS USED AS SWITCHES, R. L. Bright. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

◆

Discussion

N. L. Schmitz and **T. Bernstein** (University of Wisconsin, Madison, Wis.): Dr. Milnes' combination of gate windings and either switched transistors or rectifiers as evolved in this paper appears to have many of the advantages and few of the disadvantages of the rotating machine as a multiple-input reversible-polarity power amplifier for resistive loads. Among the disadvantages of the scheme is the duplication of control elements, namely, switched transistors in series with the gate windings of a magnetic amplifier in the load circuit.

We believe that this disadvantage is avoided with consequent reduction in physical size of the magnetic amplifiers and copper losses through the use of the circuit shown in Fig. 6. In this circuit the transistors are switched at adjustable rather than fixed firing angles by means of low-power control-circuit magnetic amplifiers. In comparing the power elements in the load circuit of Figs. 2(A), 2(B), and 3(A) of the paper with those in the load circuit of Fig. 6, it is seen that the proposed circuit replaces four transistors, four rectifiers, and four gate windings with four rectifiers and two transistors. The control circuitry is of comparable complexity in both cases. Windings *1*, *2*, *3*, and *4* are gate windings of a typical magnetic amplifier circuit rated for relatively low power service as a source of base-drive current. Bias and control windings (not shown) comparable to those shown in the paper (but on a relatively smaller size core) control the phase angle of core saturation and consequently switching angles of the two transistors.

Magnetic amplifiers have been found to be well suited to the control of transistor switching angles in single-ended nonreversible d-c power amplifiers. Preliminary studies indicate that reversible-polarity power amplification can be obtained in this manner. This scheme of operation appears to offer no greater inherent difficulty as to circulating current, loss of bias, and drift than appears in other schemes which have come to our attention thus far.

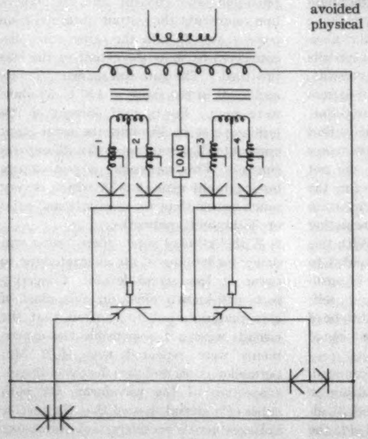


Fig. 6. Reversible-polarity d-c magnetic amplifier using transistors with variable switching angle

A. G. Milnes: The present trend from magnetic amplifier to switched transistor configurations is proceeding rapidly, and the circuits in the paper represent only a transition stage in the process. It might be said in favor of the series arrangement of reactors and transistors that transistor switching (if the load is effectively resistive) occurs at low currents, thus eliminating high peak-dissipation powers. However, in view of the success of switched-transistor d-c to a-c converter circuits this does not seem an important point, and the future of switched-transistor amplifiers, with either magnetic-amplifier-modulated drive or phase-shift control of drive as in thyratron circuits, seems promising.[1]

The comments of Professor Schmitz and Mr. Bernstein are therefore welcome and the circuit they show is particularly interesting since only two power transistors are used, whereas in the configurations that come more readily to mind four transistors are required for full-wave reversible d-c-type outputs.

Reference

1. Transistor Circuits and Applications. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 104, pt. B, Nov. 1957.

The Practice and the Economics of Applying Digital Computers to Engineering Problems

P. A. ABETTI
MEMBER AIEE

S. B. WILLIAMS
ASSOCIATE MEMBER AIEE

THE TRULY EXPLOSIVE progress achieved by the digital computer in the last 10 years, and its outstanding contributions to the solution of scientific, engineering, and data-processing problems have been recently reviewed by Harder.[1] There is no evidence that this phenomenal rate of progress will diminish in the next 10 years.

The practical application of digital computers to engineering problems has grown in parallel to the advances in the computer art. Thus the number, size, and cost of computer installations have also grown phenomenally in the last 10 years. For the General Electric Company, which makes extensive use of computers, the last published figures[2] of actual or planned medium-sized and large-sized installations are given in Table I. The great majority of these installations is used, either full time or part time, for the solution of engineering problems. It may be estimated that at present there are about 150 large-sized computer installations in the world, and about 1,000 medium-sized installations, and probably as many more being installed, planned, investigated, or considered. The economic significance of this may be assessed by realizing that the budget of a medium-sized installation may be anywhere between $100,000 to $200,000 per year, and the budget for a large-sized installation anywhere between $500,000 and $1,000,000 or more per year.

Because the art of applying digital computers to engineeeing is still in its infancy and because of the continuous rapid advances in computer hardware and programming techniques, it has been difficult to "think through" thoroughly many practical and economic aspects of planning, organizing, operating, and measuring a computer installation. Many installations, and by no means the least successful ones, have been developed more from faith than from reasoned plans supported by adequate technical or economic justifications. Others have been achieved after lengthy and detailed procedural and economic studies but in several cases, particularly in the area of application to data processing,[3] the results to date have been disappointing.

At present, there exists no established philosophy (understood as "the body of principles underlying a given branch of learning or major discipline") for the application of digital computers to engineering and data-processing work. A first attempt to establish such a philosophy for the relatively narrow field of design of electric apparatus has been made by the authors in a companion paper.[4]

Practical Questions

A philosophy is perhaps less urgent and important than a multitude of practical questions which arise continually in the mind of all groups contemplating a new computer installation, and often of groups who already have one in operation. There are specific and immediate questions, many of which must be answered before the computer is ordered or installed. For instance: What applications should be selected? What type (business or scientific) and size of computer should be selected? How much will it cost, considering both investment and running expenses? How can the economic contribution of the computer be predicted (before it is installed) or measured (after it has been in operation for some time)? What number and what type of personnel will be required? How should the procedural, programming, and operating work be organized? Should the computer installation be centralized or decentralized; should it be operated as an "open shop" or a "closed shop"? How should the communications be organized between the engineering group and the computer? What will be the effect on product quality and delivery schedules? What will be the effect on the total number of technical and nontechnical personnel? What will be the effect on the training, morale, and required abilities of this personnel? . . . and so on.

All these questions are legitimately pertinent, actual and, in many ways, controversial. It must be realized that, as the discussion becomes more concerned with specific and detailed issues, the answers are less generally applicable than a broad philosophy and must be based on the specific characteristics and circumstances of the individual application considered. Nonetheless, after these words of caution, the authors will try to answer the foregoing questions to the best of their knowledge, as based on their experience in using various types of IBM (International Business Machines

Table I. Number of Digital Computers in the General Electric Company

Type	Installed	On Order	Total
Medium-sized: *IBM 656*, Datatron, etc.	20	26	46
Large-sized: Univac, *IBM 704, 765*, etc.	8	6	14
Total	28	32	60

Paper 58-162, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 16, 1957; made available for printing November 12, 1957.

P. A. Abetti and S. B. Williams are with the General Electric Company, Pittsfield, Mass.

Corporation) computers since 1950: small, IBM *604* and the CPC (card-programmed calculator); medium, IBM *650;* and large, IBM *705*, for a great variety and number of problems. The authors' experience has been mainly in the application of computers to problems in research, development, design and engineering data processing arising from power transformers and related apparatus. However, they have followed closely the work of many of their associates in the General Electric Company in other engineering fields, and have carefully studied, through the literature or by means of personal contacts, the work done in other installations, both in this country and abroad. Therefore the following discussion will be restricted to the application of computers to engineering problems (as contrasted to strictly mathematical or to strictly data-processing problems, such as payroll), although probably some of the answers may be applicable to a broader field. Wherever possible, the authors will indicate the order of magnitude of the quantities being discussed and give examples of actual solutions to the questions posed, as drawn from their own personal experience.

Selection and Conversion

Selection of Applications

The first questions encountered in the application of digital computers to engineering problems involve: 1. the selection of specific problems to be solved, and 2. the selection of the size and type of computer. Only after these two initial decisions have been made is it possible to study the economics of the application and to make the final decision as to whether the use of the computer is technically possible and economically justified. Of course, in many cases, there may not be such wide freedom of selection. For instance, a particularly tough problem may become the bottleneck of an entire operation and must be attacked first, or the competitive situation may dictate priority of application to a certain product line in order to improve deliveries, reduce costs, or simply to "keep up with the Joneses," who already are using a computer. Also, a computer might already be available part time in the plant, for instance, in the financial section, and the cost and risk is usually less in utilizing this available computer than in acquiring another for the contemplated applications.

In general, the engineering applications to be preferred are those with high repetitiousness as a whole or in the individual portions of the programs and for which the computational and/or data-processing content is considerable. In all cases, since stored-program computers work anywhere between 100 to 10,000 times faster than the human mind, there must be sufficient volume to keep the machine busy, unless additional uses can be found for it (for instance, in the fields of accounting, payroll, inventory control, etc.), or it may be shared with other components. The requirement of high repetitiousness is dictated by the fact that, at present, programming is laborious, expensive, and time-consuming and becomes the bottleneck in most installations. Programs represent always a high investment in dollars and hard-to-get technical manpower, and only highly repetitive use of these programs ensures an adequate rate of return of this investment.

A method for evaluating the economics of a particular application will be discussed later. At present, it should be noted that in many cases it may not be advantageous to consider each application individually. This is because many applications may have certain portions of the programs in common (for instance, mathematical and editing subroutines) so that they may share jointly the total cost. Thus, many applications which would appear too expensive individually may be justified if considered jointly. Furthermore, it is very seldom in modern industry that an individual application has no relationship to the others. In many cases, the highest returns are obtained by integrating the various related applications into one single program so that the total will be larger than the sum of its parts. A typical example is the integration of design of electric apparatus with cost estimating, previously considered to be entirely separate functions, one pertaining to engineering and the other to finance. The program developed by the authors and their associates for preliminary design on the IBM *705* of power transformers (rated from 5,000 to 500,000 kva, any voltage, 2- and 3-winding) requires 90 seconds to run in total compared to the 2 weeks necessary to do the job by hand. Of this time about 70 seconds are for the engineering portion and 20 for the cost-estimating portion. It is estimated that two or three times longer running time (at several hundred dollars per hour) would be required for doing the same job with two separate programs, each with its own output and input. Thus, in selecting the applications, particularly when a large computer is being considered, it is best to consider a broad class, such as a product line (distribution or power transformers, induction motors, generators, etc.) or the major problems in a given branch of science or engineering (field problems, thermal calculations, stress analysis, etc.) rather than a multitude of isolated applications. The last case is more typical of a commercial computer service organization which usually takes whatever is offered to it by its clients and treats each problem individually.

Selection of Computer

The selection of the type and size of the computer depends on the nature of the application considered, the volume expected, and whether it will be used exclusively for engineering or for other applications also. Even if the contemplated use is exclusively for engineering, the type of application may vary considerably and influence the choice. On the one side there are the purely mathematical problems which involve a tremendous number of computations, practically no logical decisions, no data processing, a relatively small amount of input and output. Typical examples of this type of problem are the solution of a system of simultaneous linear equations, statistical multivariate correlation analyses, and the calculation of transformer reactance by the circular field method.[4]

On the other side are the engineering data-processing problems, which are encountered quite frequently in the design of electric apparatus.[3] In such problems the number of computational instructions (such as add, subtract, divide, multiply) may be 30% or less of the total. The remainder are logical decisions, reading and writing, comparing, etc., since a large portion of the design process, particularly of a standardized product, entails the reading, sorting, and selecting of already available information rather than the creation and computation of new information. In this last case, "business-type" computers with variable word lengths, high speed of logical decisions and output-input, and relatively lower calculation speed have certain advantages over the scientific, fixed-word-length, computers. For instance, in most design work, slide rule accuracy has been traditionally sufficient, and the order of magnitude of the variables involved is well known (for instance, dimensions of the components, losses per unit weight, temperature rises, etc.); thus there is no need to carry 10- or 12-decimal digits in all the computations or to use the floating decimal system. For this reason a special 5-digit "slide-rule" system was developed[6] for the Pittsfield, Mass., card-programmed calculator in 1954, in addition to the 10-digit system previously developed in Schenectady for mathematical work. This new system effectively doubled the capacity and the speed of the CPC, a small and extremely slow computer by modern standards, and allowed the performance on it of the preliminary design of the most complicated types of power transformers, such as extra-high-voltage autotransformers with three windings and 5-legged cores. Since ultimately the memory capacity, speed, and cost of computers depends upon the number of digits carried and the number of operations, there are certain advantages in using no more digits and no more operations than necessary for a specific application.

In any case a stored-program computer with the possibility of performing an unlimited number of loops, is at present required for the great majority of engineering problems. Magnetic tapes become a necessity whenever the amount of engineering data processing becomes considerable, and when it is desired to do a complete design job, not just its computations.[3]

With respect to size, there has been considerable controversy between the advocates of one general-purpose large-sized computer and the advocates of several more specialized small or medium-sized computers. In practice, both solutions are used and are operating satisfactorily, as may be seen by the figures given in Table I for the General Electric Company. Harder[1] shows very clearly that the larger the machine, the lower the cost per computation. For instance, a job costing $300 on a desk calculator would cost $1.00 on a medium-power digital computer and $0.03 on a high-power digital computer. Similarly, the time required would be 2 weeks with the desk calculator, a few minutes on the medium computer, a few tenths of a second in the large computer. In practice, as will be seen, the story is not quite as simple. Here it will be sufficient to mention that the initial investment (as contrasted to the running costs just given) is greater for the larger machines, though not proportionally to their speed. Furthermore, the ratios in speed and cost in the preparation of the input, and in some cases its introduction in the computer and the extraction of results, vary much less than the corresponding ratios just given. Nonetheless, the general statement is valid that the larger machines are more economical, provided there is enough work to keep them busy.

Another very important point, usually neglected by the advocates of several small computers, is that certain applications cannot be accomplished (economically or from any other standpoint) on the smaller machines. A simple illustration is the problem of delivering freight: if individual consignments are small, several station wagons may be better than a truck. But if some are larger than the capacity of a station wagon, it is absolutely necessary to have a truck. Thus a large computer permits applications which could not even be considered on a smaller machine, and spurs engineers to attack the toughest, most complicated and extensive problems, which are the most rewarding technically and economically and the solutions of which contribute the most to the progress of scientific and technical knowledge.

In practice, the choice of the size of the computer may be determined, at least in part, by other than strictly economical and technical considerations. For instance, if a company is decentralized in various autonomous departments, the natural tendency is for each department to want its own computer, for reasons of control, accessibility, prestige, etc. Thus the choice will be several medium-sized computers (unless a department can justify or afford a large one) rather than the joint use of one large machine in a central location. Of course, sooner or later, consideration of the economics and capacity already discussed dictates the minimum size possible. In many cases it would be absurd for each engineering unit to have its small machine, rather than one medium-sized machine for the entire engineering section. The ultimate of decentralization would be for each engineer to have his own computer, and this would mean a return to the time-honored slide-rule! However, as the progress in the computer art continues and the cost goes down, the programming becomes simpler, the trend toward smaller and less expensive machines will become more marked in many engineering organizations.

CONVERSION FROM ONE COMPUTER TO ANOTHER

Since the computer field is progressing at such a fast rate, the choice of the computer must be visualized as a dynamic rather than as a static problem. Thus the problem of selection is similar to that of shooting at a moving target. Since it takes from 1 to 2 years to plan, program, and install a medium or large machine (taking into account present delivery rates) and since the average span of a "computer generation" is of the order of 2 years, it often happens that by the time the chosen model is installed, a newer model is already on the market. Furthermore, the general experience has been that the chosen computer is loaded to capacity in a short time and a bigger and faster one is soon required to take care of the additional proposed applications. For all these reasons, the more general practice is to rent rather than to buy a computer, so that it may be easily exchanged for a larger one as soon as desirable.

The authors' experience in this respect is quite significant. In January 1953, in Schenectady they started using, part time, the CPC, which was the first commerical machine to become available. Compared to modern computers, the CPC appears completely outdated and woefully inadequate. Nonetheless, its many contributions to electrical engineering have been extremely useful. It has rendered invaluable service in introducing many engineers to automatic digital computers (as contrasted to conventional accounting machines), and as the stepping-stone to the modern high-speed, large-storage computers, either with rotating drum storage, such as the IBM *650*, or with magnetic cores and magnetic tapes such as the giants IBM *704* and IBM *705*. The experience with the Schenectady CPC was excellent, and the first power transformer was designed with it in May 1953.[6] Soon it was necessary to order a CPC exclusively for power transformer design, to be located in Pittsfield. This machine (Fig. 1) was installed in February 1954, and it was loaded to a full first shift within a month. The load increased continuously until the machine was used up to 70 hours a week. Bigger and better computers were either available, or had been announced by the manufacturers. The authors and their associates programmed some problems for the *650*, which became available in Schenectady early in 1954. While the *650* yielded very valuable results, it did not appear sufficient for the extensive scientific, design- and data-processing computations which were required in the power transformer department. At this point (May 1955) it was decided to "leap-frog" from the CPC to the *705*, omitting the intermediate step of the *650* or similar drum computers. This decision proved correct and, after 1$^1/_2$ years of extensive planning and programming work, the *705* was delivered in Pittsfield in January 1957; see Fig. 2. It was loaded up

Fig. 1. IBM card-programmed calculator, installed in Pittsfield in February 1954, and removed in June 1957

Fig. 2. IBM 705 electronic-data processing machine installed in Pittsfield in January 1957

rapidly to a full shift and it is now being used extensively for all types of computations.

One of the difficulties encountered in the frequent changeover from one computer to a bigger and better one is the rapid obsolescence of the programs, which represent an extremely heavy investment in skill, effort, and man-years of work. This obsolescence has two aspects, one obvious and one less obvious but no less important. The first is the inescapable fact that the "languages" of the CPC, the *650* and the *705* are entirely different so that the programs painstakingly prepared for one machine are completely useless for the others. The less obvious aspect is that, even if the three machines could all "talk the same language," the programs written for the smaller machines would not make efficient use of the speed and capabilities of the larger machines. In other words, it is necessary to "custom-fit" the program to the particular computer used, in order to obtain optimum utilization. This is necessary, considering that practically all computers are very expensive and are overloaded all the time, and that the economic margin between computers, accounting machines, and hand methods may be small for certain applications, particularly in the data-processing field. The discussers have found what is believed to be a workable and novel solution for this embarrassing problem. The most important and frequent applications have been reprogrammed *ex novo* in *705* language. For the less frequent applications, an automatic programming method has been devised wherein the *705* uses the same input, programs and output as the CPC, and in fact acts as a glorified CPC, but 10 to 25 times faster.[7] The ratio between bona fide *705* programs and CPC is betweeen 60 or 80 to 1, so that considerable speed is being sacrificed by this artifice. However, approximately 25 man-years of programming and "debugging" were saved in this manner, and thus this expedient is definitely economically justified for the less important applications. Having thus made the CPC completely obsolete, it was returned on June 1, 1957. When a conversion system is not available, there is a tendency to retain the older, less effective computers, and this is generally uneconomical and contrary to the natural tendency of the engineer, which is to use always the most advanced techniques available.

Organization

PROGRAMMING AND OPERATION

Before discussing the organizational and economic aspects of a computer installation, the various steps involved in actually carrying out the application of a given important engineering problem on a computer will be described. This step-by-step process is described for a large or medium sized machine with magnetic tapes, which are almost indispensable for any extensive amount of engineering data-processing work. For less important problems, or for a small computer without tapes, the process may be less formalized, and some steps may be omitted but, in any case, the description is still generally valid. First, one must distinguish between two distinct phases of the problem 1. preparing the program, and 2. running the problem on the computer whenever required for production (as contrasted to "debugging"). These are, conceptually and practically, two different kinds of work, and in many cases may actually correspond to different organizational components.

Preparing the Program

1. Meetings are held between the engineers interested in the particular application and people assigned to the computer to define the problem, establish the objectives, and generally discuss the method to be followed. These preliminary discussions may require days, weeks, or months, especially for design synthesis applications.[3]

2. Engineering procedures and programming specialists analyze the problem and prepare flow charts (or block diagrams) showing the flow of information, the logical decisions, and the computations to be made, etc. This is the first step in formulating the program, which is a series of computer instructions. Flow charts may include any number from a few to several thousand blocks.

3. Programmers and coders code the flow chart, i.e., they translate the instructions into computer language. Engineering programs may have any number between several hundred to 100,000 separate computer instructions.

4. The coded instructions sheets are keypunched into program decks of punched cards. These may include any number of cards from a few to several thousand.

5. The program is assembled into its final form by the machine itself and then is tested or debugged by the programmers, who run test cases on the machine and correct errors as they appear. This process may take from 1 to 100 machine hours, and from a few days to months of elapsed time. One hour of machine time is required to debug several hundred program instructions, the number depending on the complexity of the program and in the ability of the programmer.

6. Forms for input data are prepared.

7. The input and output of the test runs is reviewed with the interested engineers, and further corrections and refinements are made.

8. The final checked program is transferred from cards to magnetic tape and is now ready for the production runs. One or more tapes may be required, depending upon the application.

9. The programs are modified and updated as the need arises, upon suggestions of either the engineering group or the computer group.

In practice, many of these steps merge into each other and they may be repeated several times. Also, there is a continuous feedback from one step to the preceding one. In checking out a program it is necessary to ascertain that: 1. the instructions are valid with respect to the particular machine code being used; this is relatively simple, and most errors are of a trivial nature, and 2. the program, logically and computationally, achieves the desired objectives for all possible variations to be encountered in practice. This stage is quite difficult, and requires a thorough knowledge of the problem and the computation process, as well as continuous feedback from the interested engineers and managers.

Running a Job for Production

1. The engineer assigned to a job fills in the input forms according to the specifications and any other special requirements. One or more sheets may be required.
2. These forms are key-punched by the computer operations group onto punched cards. The number of cards may vary from 2 to 10 or 20.
3. Input data cards are sorted and checked on auxiliary equipment for possible errors.
4. If there are many jobs and many cards, these cards may be transferred to magnetic tape by auxiliary card-to-tape operation, to save running time.
5. Following the program instructions, the computer takes the input data, performs the required calculations, and prints the output directly, or writes it on magnetic tape, if it comprises several sheets.
6. By auxiliary tape-to-printer operation, the output is printed out on the standard forms.
7. The output is transmitted to the engineer responsible for the problem. In case of error, e.g., incorrect input data, the machine "hangs up" and he is told what should be corrected.

This running procedure may appear quite detailed but it is designed to save as much as possible of the computer time, which is worth from less than a hundred to several hundred dollars per hour. It should be realized that jobs which would take weeks by hand require only a few minutes on the computer. Therefore the elapsed time between submission of the input data sheets to the computer group and the transmission of the results to the engineering group may involve only a few hours whenever the procedure and the required communications are properly organized. For instance, the Pittsfield *705* is reserved according to a planned regular schedule, for processing, every working day, daily design requirements such as propositions, requisitions, and special studies. If the input data sheets are given to the computer operations component early in the morning or early in the afternoon, the results will be returned to the design office before the end of the morning or of the afternoon respectively. Thus, the total elapsed time is less than 4 hours, corresponding to weeks of work by hand. Other less urgent engineering requirements, including debugging of new programs, are taken care of daily as time is available, with a less rigorous schedule.

ORGANIZATIONAL PATTERNS

The foregoing description of how a design application is programmed and run repeatedly on a computer leads to a discussion of organizational questions. There is at present a great variety of organizational forms, depending upon the size and type of computer, the type and frequency of applications, the time the computer has been in use, and the company's organizational philosophy. In the following paragraphs are presented the authors' ideas on this subject, and how they have been embodied in practice in the Pittsfield *705* installation.

In the preceding discussion three different groups have been recognized: 1. the engineers, charged with the responsibility of obtaining solutions to various problems, either by hand or by computer; 2. the programming group, responsible for preparing correct and economic programs for the selected applications; and 3. the computer operations group, responsible for running the requested individual jobs on the computer correctly, efficiently, and on schedule. In several practical cases, particularly in the smaller computer organizations, or in the initial stage, these three groups may merge into each other and the same persons may be charged with more than one of the responsibilities listed. Nonetheless, it is important to recognize that there are three different kinds of work, at least conceptually. As soon as the computer is loaded up it becomes necessary to identify and split off from their group one or more persons responsible for actually running the machine, allocating time, key-punching the input, etc. This is the nucleus of the computer operations group, which may comprise any number from 1 to 10 or more persons. If the computer is not in the engineering component, e.g., if it is in the financial component, this identification of the operations group is obvious.

The separation between engineering and programming work may be less apparent and less immediate. In fact, there are two schools of thought concerning engineering and scientific applications of computers. In the "open shop" all interested engineers are given programming courses and, in some cases, even courses in operating the machine. The engineers are then free to prepare their own programs, debug them, and run them for production as they see fit and whenever there is machine time available. In the "closed shop" all programming is done exclusively by "professional" programmers in a separate component and they alone are allowed to prepare and debug programs. Thus the engineers have no direct contact with the computer. All intermediate variations are possible within these two extreme steps. For instance, in the "swinging-door shop" there is a special group mostly comprising engineering technicians, who have a thorough knowledge of the computer. The engineers direct the programming work, but utilize this group for coding, debugging, and operating details. In other variations, one or more programming experts are available, to advise the engineers who do their own programming when these encounter difficulties or otherwise require help in the various stages of their work.

Returning now to the two extreme cases, the closed shop and the open shop, it is clear that both have their advantages and disadvantages. The open shop has all the well-known advantages of decentralization. The engineers are responsible for the entire process; there are no barriers of working knowledge or organization between them and the computer; they have knowledge of the entire process and thus consider the computer as an extension of their knowledge and as an adjunct or helper rather than as a little-known stranger who threatens to take away their job. Individual characteristics of each engineer are reflected in the programs and there is less of a shock in the transition from the traditional design methods to the new computer methods. This approach also avoids the bugaboo, more imagined than real, of concentrating all the knowledge and decision-making in a small group of "experts." Probably this type of organization is best for small and medium-sized computers, in the initial stages, and where there is a great variety of jobs which are not run regularly and often, on the computer.

The closed shop also presents many advantages. As mentioned, programming represents in any computer installation a considerable, if not the most important, investment of time, money, and effort.

This is particularly true for larger machines, and when the data-processing content of the job is high. Therefore management will want the investment kept within reasonable limits, and to yield a high rate of return. Also, a new computer is often loaded up to capacity within a relatively short time, and prospective applications clamor for more time. This means going to another shift or obtaining another machine, with a corresponding increase in rental and operating cost. In short, the circumstances invariably lead toward the most efficient use of the programs and of the available computer time. It is a fact that two identical applications (i.e., those having the same input and output) may correspond to programs of widely different length and complexity and to running times of ratios as high as three to one. Thus the total cost of an application may vary very considerably, depending on the ability and training of the programmers, the organization, the pattern of communications, etc. In many cases this may spell the difference between the success and the failure of the application, technically and economically.

For complex, expensive, and highly repetitive applications, there is therefore considerable merit in having a group of professional programmers, thoroughly trained in the use of the machine being used, and assigned full time to a given class of applications. It takes at least 3 months to train thoroughly a programmer, and even more time to become an expert in the procedural and communication aspects of any application. In line with the present tendency of science and engineering towards more marked specialization, all programming and debugging is entrusted to a group of professional specialists, programmers, coders, and operators. Among these may be found specialists in input and output forms, use of auxiliary equipment, automatic programming, numerical methods, cost estimating procedures, etc. It has been stated repeatedly that 10 to 20 persons are required to "keep busy" a medium sized computer, and 30 to 60 persons a large one. Problems of planning, organizing, integrating, and measuring the activities of such a large group of highly specialized persons make it almost mandatory to have a separate organization with its own authority, responsibility, and accountability. Such an organization is also more attractive in hiring and keeping programmers, which are at present very scarce and relatively expensive, because it provides the feeling of togetherness, the possibility of self-development, and more opportunities for promotion. Also, the contacts and communications between the programming and operations group, vital for the success of the application, may be better formalized, and conducted more efficiently. For all these and similar reasons, a separate programming component appears preferable for a large installation, for complex and difficult applications, and for those with high data-processing content.

In the closed shop there is the danger, real or imagined, of concentrating all the working knowledge and decision making in this group and, in fact, of creating a duplicate of the engineering office, with ensuing rivalries, jealousies, and "empire-building" tendencies. Of course, this is the same problem encountered in every line-staff relationship, and management can very well take care of it. The programming group must be reminded that it is strictly a service organization and that the responsibility of designing the product still remains entirely in the design component. It is also necessary to maintain continuous and effective formal and informal communication between engineers and programmers. In many cases, transfer of personnel between the two groups is an excellent source of cross-fertilization of ideas, new ideas, and better ways of doing things. Generally speaking, it is easier to teach programming to an experienced engineer than to teach a particular branch of engineering, such as transformer design or stress analysis, to an experienced programmer. (Besides, the latter are in short supply, and employing them is costly.) Thus, in most cases, the programming group will consist of several engineers, among them some of the most experienced and able, and there should be no danger of the work of the two groups "going off in different directions." The programmers, however, should not be obtained exclusively from within the engineering office. Some new blood is desirable, in the form of engineers or mathematicians employed from outside the company, to bring in new ideas and techniques, and the experience gained from work with other computer installations. Probably the best ratio is of two programmers transferred from other groups in the plant to every one hired from outside.

Purposely omitted, up to this point, has been any discussion of the reporting organization of the various components, that is, whether they should be part of engineering, finance, general administration, or of a separate *ad hoc* computer component, reporting to the department or plant manager. We feel that this much-discussed and still controversial aspect of the problem is less important than the recognition that there are three different kinds of work. In practice, all of the aforementioned solutions in all their possible combinations have been tried. For instance, in some installations engineering is responsible for programming and running the computer, in addition to its usual duties. In other cases, the financial section is responsible for programming and running the engineering applications and it has several mathematicians assigned to its computer groups. Many other organizational patterns exist, and they apparently operate quite sucessfully.

THE PITTSFIELD CPC AND IBM-705 INSTALLATIONS

To summarize the authors' experience in Pittsfield, at first the CPC was assigned to the advanced development component of the power transformer department and operated more or less as a closed shop, three engineers being responsible for all the programming. The responsibility of operation was shared between the two computer operators and the engineers themselves. As soon as the principal programs were completed and performing satisfactorily, the CPC was transferred to the design component and was operated henceforth mostly on the open-shop basis. When it was decided to acquire a *705* and create a large group of professional programmers, a separate component was set up. All programming was done by this component strictly on the closed-shop basis, using the only two systems then available from IBM, the symbolic and the autocoder. Shortly before the *705* was delivered, the computer component was reorganized into two separate units, one responsible for programming as previously, the other for computer operations. As soon as a program is completely debugged and tested successfully for several practical cases, it is turned over by the programmers to the operations component. From then on, the operations people deal directly with the engineering office and are alone responsible for running the jobs. The programmers are consulted only occasionally, in case of particularly difficult jobs, errors, doubtful results, etc., but essentially they have nothing to do on these programs (except for updating) and are free to start working on other applications.

At about the same time, two other types of programming methods became available. One was the CPC programming system, converted automatically to the *705*, as discussed previously.[7] This is very important because there were more than 50 engineers and technicians in Pittsfield who knew how to program for the CPC, but not for the *705*. Now, by this method, they can make use of the *705* without any further training. Early in 1957, IBM made available a new automatic programming system for the *705*, called "Print I" and especially designed for scientific and engineering applications. While the symbolic *705* coding system requires a 4-week course, practically full time, the Print I system may be taught in a week to technical personnel. It may be learned in an even shorter time by those already familiar with CPC or *650* programming. Compared to ordinary *705* programs, Print programs require more time for assembly and for running, since all operations must be interpreted and all calculations are done with 10-digit floating decimal numbers. As has been seen, for design work, four to five significant digits are usually sufficient, and fixed decimal is entirely adequate, since the order of magnitude and range of the various design quantities is known. Thus, machine running time, which in the *705* is proportional to the number of digits carried and to the number of digit shifts, is saved by using ordinary fixed decimal code and variable word length. Nonetheless, Print I is extremely useful and fast for special design and development problems which are not run very often or which are relatively short. Therefore, Print I courses were given to many engineers and technicians who are not part of the computer component. They can now use the *705* on an open-shop basis for any program written in Print I language. An experienced specialist in automatic programming is available within the computer programming group, to advise and help anyone using the Print I system.

We feel that in this manner, the Pittsfield *705* installation combines the high efficiency and professional ability of the closed-shop organizations, for the more complex, lengthy, and highly repetitive programs, with the accessibility and decentralization and spreading of responsibility of the open-shop organizations for the other programs. So far, this system has been working well, and it has met with the approval of both the engineering and programming groups.

Economics

There is no doubt that the most vital questions and the most frequent ones asked by management concerning a prospective or an existing computer installation are, "How much money will it save?" or "How much money is it saving?" In a free-enterprise society, wherein profits and losses are one of the principal indexes of measurement, this could not be otherwise. On the other hand, many difficulties are encountered in assessing the actual dollars and cents saved by a given installation, even after the first year of operation. More difficult is the question of evaluating the savings of a prospective computer application.

First of all, it should be recognized that not all computers are installed to save money. In certain cases, a computer is run as a laboratory devoted to pure research. The scientists are given the tools with which to work in a certain field in the laboratory without any immediate goal of financial returns. Similarly, a computer may be obtained by management so that its scientists, engineers, and financial personnel may experiment with it with a view to finding later good use for it. As is well known, many of the most important advances have arisen from this kind of approach and from the principle of serendipity, and therefore it should not be dismissed too lightly.

In most practical cases, however, the computer is competing, in dollars and in manpower, with other alternate investments, and therefore the decision to obtain or to keep one must be justified economically. The dollars-and-cents aspect may be very important for certain applications, particularly in the data-processing field (for instance, payroll and tax accounting) but very much less so for other applications (for instance, in the advancement of new technologies vital to the national defense). Therefore, the discussion of economics will not be exclusively of dollar savings, but also savings in time and effort, and of other contributions usually defined as "intangibles" by accountants.

The application of digital computers to engineering problems may therefore be justified, for one or more of the following reasons: 1. It saves money. 2. It saves time. 3. It saves the effort of skilled personnel, of which there is at present a shortage. 4. It improves accuracy and the quality of the product. 5. It enables engineers to solve problems which could not be solved without it.

DIFFICULTIES IN DETERMINING SAVINGS

The normal accounting practice in determining the savings which may be attributed to a specific innovation can take into account only reason 1, dismissing the others as "intangibles." Even in the determination of the actual dollars and cents saved, the two alternate methods, by hand and by computer, must be directly comparable, i.e., starting from the same input data, they must produce the same output data, after carrying out the same computations and logical decisions. This may be the case for a standard payroll application, which starts from the employee's clock cards and produces pay checks, and various accounting reports. It may also be the case for certain relatively simple and repetitive mathematical problems, for instance, the solution of four or five simultaneous linear equations. An engineering assistant with a desk calculator will require so many hours, a computer so many seconds, and thus a direct comparison is possible if the total costs per unit of time are known. In the great majority of cases, however, engineers want to use the computer not just to repeat narrowly and uncreatively what was being done previously by hand, but rather to obtain the best possible result, with regard to accuracy, quality, and speed, or to discover solutions to previously unsolved problems. This, of course, is the right approach and the only one conducive to technological and scientific progress. Even if the two alternatives were exactly the same in procedure and quality, they might not be the same in quantity. For instance, with the computer readily available, engineers will usually solve many more systems of linear equations than previously, in order to try out more alternatives of the same problem. Since, strictly speaking, one can only "save" the money that he was spending previously, it is hard for the accountant to assess the "savings" corresponding to work which was not done previously.

Even assuming that the computer does exactly the same work, in quality and quantity as previously done by other means, it may be difficult to evaluate the actual dollars saved. This will be made clear by the following hypothetical example. In three identical engineering sections, consisting each of 100 employees, 70 were employed for production design work and 30 for development work. After the introduction of a computer to do design work, only 30 employees are now necessary to complete the same amount of production. Furthermore,

1. In the first section, the same amount of production is now being done with 30 employees. The other 40 were released for other work outside the section.

2. In the second section, 40 production

design people were shifted to more highly paid jobs in development within the section.

3. In the third section, no changes were made in personnel, but the number of designs produced by the production group of 70 people was more than doubled. These additional designs are used to take care of the increase of business, to reduce the amount of material and labor by considering the best alternatives, to improve service on customer applications, etc.

It is clear that only in case 1 have actual savings been made in dollars and cents for that section. In case 3 the savings are no less real, but are for the entire department in reduced manufacturing cost and in increased business. In case 2 the budget of the section has actually gone up in order to obtain increased profits in the future by stepping up the development effort. The important point is that in all three cases the "gross savings" attributable to the computer are exactly the same, namely, 40 production people. In dollars this would correspond to the total salaries plus overhead. The contribution of the computer is totally unassociated with the decision of how and when to use the 40 people it has displaced. This decision is strictly the responsibility of the management of the section involved.

In addition to the savings just discussed which may be attributed legitimately to the computer, there are other savings, no less important, which are achieved concurrently to the introduction of the computer. Both in data processing and in engineering, these savings arise from improved practices and procedures, standardization, elimination of useless forms and reports, etc. It has been stated several times, half seriously and half facetiously, that every company should order a computer, prepare for it, and then cancel the order. In this manner, considerable savings would be realized, through improvement and streamlining of existing obsolete procedures. This statement applies also to engineering, particularly to design, and up to half the savings obtained with the use of computers may be attributed to improved methods and standardization; the remainder to the computer itself. The point, however, is that none of the savings were actually realized without the computer, which acted as the catalyzer of the entire process. After all, management is interested in the total savings obtained, and it is a moot question as to how much should be attributed to procedure improvement by itself, how much to the computer by itself, and how much to the interaction of these two factors.

THE "GROSS CONTRIBUTION"

The foregoing discussion has shown that the determination of the potential or actual savings attributable to a computer is far from simple, even if one is restrained from placing a dollar tag on intangibles like time, service, quality, accuracy, more complete and quicker information, etc. In conclusion, the discussion will be of the "gross contribution" or "net contribution" attributable to the computer, rather than of gross savings or net savings. The latter is strictly an accounting problem, and undoubtedly each company will solve it according to its established practices. Therefore the authors believe that the gross contribution attributable to a given computer should be figured out on the basis of the cost of doing with the tools previously available the same jobs in quantity and quality as are now done on the computer. In this manner the gross contribution may be evaluated with respect to hand methods, conventional punched-card equipment, and other types of computers (usually smaller ones). The net contribution is obtained by subtracting the total computer cost from the gross contribution. These costs will be discussed in the following section.

COMPUTER COSTS

The over-all costs encountered in the application of digital computers to engineering problems may be divided into two categories: 1. investment costs, consisting of the sum of programming costs (including debugging), physical installation costs, and starting-up costs (such as running in parallel for a certain period, conversion of basic data to punched cards and magnetic tape) etc.; and 2. running expense, which includes computer rental, salaries of operating personnel, supplies, and overhead. These two categories correspond to the two different kinds of work discussed in the section entitled "Organization."

Scientific programming cost may vary between $1.00 to several dollars per computer instructions, as quoted in the literature. The higher figures are for complex data-processing applications, and for the larger computers. Of the total amount, approximately one third represents debugging costs. Since an engineering program may have between several hundred to several tens of thousands steps, the programming costs may be any amount from several hundred dollars to $100,000 or more. Even higher amounts are possible for data-processing applications, and figures from $4.00 to $10.00 per computer instruction have been quoted in the literature. The cost of the physical computer installation includes the costs of preparing the computer room, the auxiliary punched-card equipment room, the programming office, the power supply, the air conditioning, etc. They may vary from $10,000 for a small installation to several hundred thousands for a large one. The prorating of these costs to the various jobs done on the computer may be done in various ways; it is essentially an accounting problem. One way is to prorate the cost according to total running time of each job per year. The start-up costs are usually much smaller for engineering than for data-processing applications. They probably are worth considering only for engineering data-processing applications such as design synthesis of complex apparatus.

Running expense, for one shift, will be any amount from less than $100,000 per year for a medium-sized machine without tapes to $600,000 or more for a large machine. As noted, the running cost per operation is less, the larger the machine, since the speed increases much faster than the cost.

As has been mentioned, total costs for a given application may vary widely depending upon the ability of the programmers and, in a lesser degree, upon the computer selected and the ability of the operators. It is also possible to trade programming costs for running expense. For instance, a "tight" program with many loops is more difficult to program and to debug, but it takes less time to run than a "loose" program. This consideration is at the basis of many automatic interpretative programming methods, wherein ease and speed of programming is paid in part by longer running times. Similarly, it is possible to use for the same group of applications either one general program which will process all possible combinations of the input; or several special programs, each for a particular application. The smaller the computer, in general, the less general the program. Since loading time may be quite expensive for a large computer, a general program is usually cheaper to run than several separate programs. A more general program, on the other hand, is more difficult and laborious to write and debug, since its various portions are interrelated by loops and its complexity increases faster than the number of program instructions. For preliminary design of transformers on the CPC, different program decks were used for different types of components,

Vol.
Pt. 1
1958

e.g., windings. These decks were selected and collated manually by the operator. In the *705* there is only one program for the preliminary design of all power transformers, regardless of rating, number of windings, etc.

In conclusion, over-all computer costs are very much dependent upon the ability of the programmers and operators, who must adapt their techniques to the specific characteristics of the application and of the computer being used. Having thus discussed methods for evaluation of the gross contribution of the computer and of the over-all costs, these methods will now be used to determine the net contribution of a specific application in power transformer design. Relative contributions and costs will be used rather than actual dollars, because the actual costs vary so much between different intallations and, furthermore, are not usually available. Therefore it is felt that relative values will be more significant and more broadly applicable for comparison with the economics of other engineering applications.

TRANSFORMER REACTANCE CALCULATION

A typical example of an engineering application of small-scale and large-scale digital computers is the calculation of reactance between any pair of concentric transformer windings. Several of these calculations must be performed for each transformer design, and the specified impedances must be met in test within relatively narrow limits. In brief, conventional methods of calculation by hand have always assumed that the windings are flat, without curvature. This, of course, is not the case in practice, and this assumption loses its validity when the ratio of the gap between windings to the average radius is large, as, for instance, in multiwinding transformers, or when the windings are axially nonuniform. The errors introduced from this cause may occasionally result in rejection of a unit after it is built, when the impedance test shows that the guarantees have not been met. In any case, this limitation requires higher design safety factors, and closer manufacturing tolerances which, in turn, mean higher cost. While this problem had been recognized for many years, it always appeared impractical, if not impossible to introduce curvature into the reactance calculations by hand, because of the Fourier series and the Bessel asymptotic series involved, all with a large number of terms.

After the CPC was installed in Pittsfield, a program was developed for calculating reactance by the circular field method between any two transformer windings.[4] An extensive comparison of calculated and measured values showed that the accuracy of this method was considerably higher than the accuracy of the hand method. Since the CPC was loaded to capacity, and because 80 minutes were required for one reactance calculation (compared to a minimum of 60 engineering hours by hand) only the most important designs were checked by this system. When the *705* was installed, a new fixed-decimal program was prepared for this computation, requiring only 57 seconds to run. Table II shows a comparison of times and relative costs for the three methods: by hand (which was never used in practice), by CPC, and by *705*. It is evident that the time and cost of this accurate method, using the *705*, is insignificant. In fact, less time and lower cost were involved than in doing the same calculation by the approximate method, by hand. Therefore, reactances of all power transformers are now computed by *705*, thus ensuring the highest accuracy in all cases, and eliminating totally the eventuality of rejection in tests. It is interesting to note from Table II that, while the CPC achieved a very important reduction in cost with respect to hand methods (89%), the introduction of the larger computer achieved considerable additional savings (from 89% to 98.7%). More important, it was only the extremely short-running time on the large computer that made the adoption of this method possible for all power transformers, rather than for only a selected few.

Table II. Comparison of Times and Costs for a Reactance Calculation

	By Hand	By CPC	By 705
Time	60 hours	80 minutes	57 seconds
Relative time	1	1/45	1/3,800
Relative cost per hour	1	5	50
Relative cost per calculation	100%	11%	1.3%

The *705* program has approximately 2,000 computer instructions and it required about 3 man-months to program and 4 machine-hours to assemble, debug, and check out completely. It may be estimated that the ratio of the investment for this program to the running costs for one computation on the *705* is approximately 1,200 to one. From the figures given in Table II, it may be concluded that, with respect to the hand method, the investment will be paid off after 16 computations, and after that the net contribution, of the computer will amount to 98.7% of the cost of one computation by hand. With respect to the CPC, the corresponding figures are 160 computations and 88% of the cost of one computation by CPC. Since several reactance computations are being made each day on the *705*, the period required for paying off the investment, with respect to the CPC, is of the order of 2 months, less time than was needed to write the program. It may be concluded that, even considering the total costs, the *705* is considerably cheaper than the CPC for this application.

It should be stated that this reactance program is only a small fraction of the engineering programs now available for the Pittsfield *705*. These programs comprise well over 100,000 instructions. The magnitude of the total job may thus be evaluated in terms of man-years and dollars expended. As an example of the results obtained, a final design of a large power transformer, requiring weeks by hand, is now done completely in a few minutes on the *705*. In many cases the contribution obtainable from the computer as a result of design optimization[8] is probably considerably more important than the manpower savings discussed in the foregoing.

Consider now some of the intangibles.

TIME

As already mentioned above, design calculations are usually processed in less than four hours in the *705*, while they would require weeks by hand. This collapsing of time is very valuable in design work, because it is possible to give better service to customers' inquiries and faster deliveries of their orders. In development work, it is possible to evaluate much faster the results of tests, and to decide immediately whether more tests are required. It is also possible to determine much faster the effect of new technological developments or new ideas on the product. Thus the computer makes available much sooner to the customer the results of research and development in the form of an improved product.

ACCURACY AND QUALITY

As has been seen, the reactance computations done on the computer are much more accurate than those done by previous methods. This is also true for many other design applications. This type of contribution, though difficult to measure in dollars and cents, is a very important addition to the net contribution. One possible way of measuring this advantage is to consider the savings effected by the avoidance of rejections in tests, and the savings obtainable by reduction of the design safety factors. If it is decided not to change the safety factors, the contribution is in effect passed on to the customer, in the form of a better product, more liberally designed. Even after evaluation in dollars of these two contributions, there remain the intangibles, such as increased customer good will as a result of higher quality and better service. These advantages, with their long-range profits, will in turn influence favorably the company's position in the mind of the public.

EFFORT

Even if no savings could be attributed to the computer, it would still save time and effort. With the present shortage of technical personnel, now vitally needed for national defense, it is important to utilize the available manpower as efficiently as possible. In effect, the computer frees engineers, mathematicians, and technicians from the effort of routine calculations and makes their time available for truly creative work. The long-range effect on the national economy of this particular contribution of the computer cannot be underestimated. Specifically, in the field of electric power, the load is expected to double every 10 years. This increase would require a corresponding increase in technical manpower of the utilities, manufacturers, research institutions, etc. Since the number of electrical engineers expected to graduate in the next 10 years will certainly not double, the computer will be one of the most effective means for decreasing the gap between requirements and availability.

Effect on Personnel

It is certain that computers will not cause any significant displacement of technical personnel in the foreseeable future. Even for the most typical data-processing applications, for instance, in insurance companies, the impact of office automation on personnel is much less pronounced than was feared or expected. Yet some of these applications are often made for the specific purpose of reducing clerical costs. On the contrary, it may be stated that computers are used in engineering not so much to reduce the number of engineers and technicians, but rather to alleviate their shortage, improve quality and service, reduce time, and take care of the increase in business.

Perhaps more interesting is the effect on the type of personnel which will be required in a future organization that makes extensive use of computing facilities. Anticipating this is somewhat like gazing into a crystal ball, but still worth considering.

It may be assumed that, for conventional methods, the distribution of the required skills follows a Gaussian distribution curve. The effect of the computer on the personnel required would appear to reduce the frequency of persons with average skill and, on the other hand, increase the frequency of persons with higher and lower skills. In fact, the skill required by the procedure and programming specialists is higher than average, and the emphasis will be on increasing the development effort to improve the quality of the product. Since the computer can produce results faster and in much greater quantity than previously, higher ability will be required to keep up with its pace and make optimum use of the new information available. This, of course, is a restatement of the familiar fact that a machine is no better than the persons who are using it. Since computers represent high initial investments and yearly expenses, it is necessary to employ high-quality personnel in order to obtain adequate returns. On the other hand, the processing of data by means of the computer can be done by young engineers and technicians, with less skill than that of the average engineer. This trend should be valuable in attracting more engineers toward more interesting, challenging, and higher-paid work, and in utilizing technicians to release engineers for jobs which require the complete application of their skills. The training of young engineers is facilitated and accelerated by the use of the computer. Consider, for instance, the training of a designer: once he has mastered the fundamentals, he can make many more designs on the computer than by hand in the same period and he can observe more readily the effect of changing the specifications or the design variables, and compare alternate forms of constructions.

Since engineers are normally progressive individuals, eager to use the best available tools, the effect on the morale of the engineering force of installing a computer will be good, provided it is explained thoroughly to all engineers before its installation. It must be made clear that it is a willing and docile helper, an extension of their skills, rather than a mysterious black box with push-buttons which threatens to take away their jobs. Programming courses, either mandatory or with voluntary participation, are also very effective in removing the aura of mystery and misunderstanding which often accompanies the word "automation," and in spurring all engineers to make optimum use of the new facility.

Summary and Conclusions

The foregoing is an endeavor to analyze the most important problems facing all persons planning or managing a computer installation for engineering work. The selection of the applications to be programmed, the selection of the computer, and the problems of conversion from one computer to another are discussed. The various steps for programming and running an important engineering problem, and the organizational patterns which arise from these activities are described. The economics of computer installations are discussed at length and it is shown that the conventional accounting methods are inadequate for evaluating the over-all contribution of the computer. Finally, the effect on the number, quality, and morale of personnel is considered. This analysis is supplemented by actual examples taken from the authors' experience.

The discussion leads to the following conclusions:

1. Applications should be selected which have a high degree of repetitiousness. They should be planned and evaluated as a group, rather than individually.

2. The selection of the type and size of computer depends upon the applications considered, the amount of usage, and organizational considerations. Larger and more expensive computers are required as the volume and the amount of engineering data processing increases.

3. Conversion from one computer to a larger and more economical computer may sometimes be made simpler, faster, and cheaper by devising automatic programming techniques to simulate the work of the smaller computer on the larger.

4. Various patterns of organization are possible, depending upon the circumstances. In any case, three kinds of work may be identified: engineering work, programming work, and computer operations. A closed shop appears more advantageous for large-scale machines and for highly repetitive applications with large data-processing content, such as the design of electric apparatus.

5. Since the engineering work done on a computer is seldom directly comparable to the work previously done by hand, it is difficult to determine savings by ordinary accounting methods. A method is presented for determining the computer costs, the gross contribution, and the net contribution attributable to the computer.

6. In most cases the contribution of the computer in the areas of optimization, reduced time, increased accuracy and quality, and the freeing of technical manpower

for creative effort, is more important than the savings obtainable through reduction of technical personnel.

7. The effect of computers on the number of engineering personnel will be relatively small. More important is the effect on the quality of technical personnel which will be required in the future, on their training, and on their morale.

8. The over-all contributions to date of digital computers to the advancement of science and engineering are significant, technically and economically. It is felt, however, that engineers are just starting to learn how to apply this powerful innovation and that future contributions will be even more important.

9. A thorough understanding of the theory and practice of applying digital computers to engineering problems will be necessary to achieve optimum utilization of this revolutionary new tool.

References

1. The Computing Revolution, **E. L. Harder.** *Electrical Engineering*, vol. 76, June 1957, pp. 476–81; July 1957, pp. 586–90.

2. Never Overestimate the Power of a Computer, **R. F. Lewis.** *Harvard Business Review*, Boston, Mass., vol. 35, 1957, pp. 77–84.

3. Philosophy of Applying Digital Computers to the Design of Electrical Apparatus, **P. A. Abetti, W. J. Cuthbertson, S. B. Williams.** *AIEE Transactions*, vol. 77, pt. I, July 1958, pp. 367–79.

4. Transformer Reactance Calculations with Digital Computers, **L. Rabins.** *Ibid.*, vol. 75, pt. I, July 1956, pp. 261–67.

5. A New Five-Digit Plugboard System for Card-Programmed Calculator, **S. B. Williams, Miss N. M. Noonan.** *Ibid.*, Mar. 1956, pp. 67–71.

6. Application of Digital Computers to Transformer Design, **S. B. Williams, P. A. Abetti, E. F. Magnusson.** *Ibid.*, pt. III, Aug. 1956, pp. 728–35.

7. A Simulation of the Card-Programmed Calculator on the IBM 705, **S. G. Fleming.** *CP 58-230* (available on request).

◆

Discussion

H. R. J. Grosch (International Business Machines Corporation, New York, N. Y.): A useful rule for forecasting computer savings is that "economy is as the square root of the speed."[1] That is, to reduce the cost of a calculation by a factor of 10, one must utilize speeds faster by a factor of 100.

The authors mention, as certain advantages of the "closed shop," professionally expert programming, more efficient operation of the computer, and opportunities for development of advancement of personnel. The list may be extended to include:

1. Interest in improved equipment, and in more advanced methods of programming currently installed equipment.

2. Attendance at specialized meetings, membership in specialized societies, and subscriptions to specialized publications.

3. Contributions to the profession.

4. A center which can advocate computer applications in other areas such as accounting, production control, sales analyses, operations simulation.

Establishing the contribution of the computer operation to the business, in order to obtain or increase financial support, is described as a major task. One method which I adopted in several large installations alleviates the difficulty: services are sold to various technical projects and functions, on an hourly basis. Rates for machine time and for programming personnel are established so as to liquidate expenses at the predicted utilization rate for the budget period. As the authors indicate, utilization often builds up with unexpected rapidity; rates may be lowered or rebates given, or the surplus used for advanced personnel training, for support of programming research, or to experiment with new areas of application not yet able to pay their way.

The burden of proving the advisability of using the computer group is spread among project and functional managers. Budget problems are minimized, being reduced largely to collecting the estimates of prospective users and multiplying the sum by an experience factor to allow for the new work always to be expected. Multiple-shift operation permits a very flexible capacity, and this flexibility may be increased by utilizing service bureaus and programming subcontractors; thus, proceeding for a year or two while experience accumulates need not be very difficult.

Reference

1. High Speed Arithmetic: The Digital Computer as a Research Tool, **H. R. J. Grosch.** *Journal*, Optical Society of America, New York, N. Y., vol. 43, Apr. 1953, p. 306.

C. G. Veinott (Reliance Electric and Engineering Company, Cleveland, Ohio): I agree qualitatively with a large part of this interesting paper and with most of the authors' conclusions. However, I believe that the economics as they see them are not necessarily applicable to the needs of a smaller engineering organization; it is with such cases that the following remarks are principally concerned.

In general, we agree with the authors and with Harder that the cost per computation is less on a large-sized expensive computer than on a small one. This, of course, assumes that there is sufficient work load to keep the large-scale computer more or less continuously busy. It seems to me that the most economical computer is the smallest one that will just handle the work load adequately. Of course, the work load normally increases rapidly and some anticipation of future needs must be made. The authors refer to the IBM *650* as a medium-sized computer which they report as having an annual budget of about $100,000 or upward. Excluding the operating personnel, the machine rental of an IBM *650* together with associated equipment is approximately $5,000 per month. The monthly rental on an Alwac III with a 4,000-word drum, including maintenance, is approximately $2,000 per month, or less than half the rental cost of a *650*. With an 8,000-word drum, the Alwac rental is only about $2,200 per month. The computing speeds are roughly comparable, or are at least of the same magnitude. Carr and Perlis have compared the speeds of these two computers,[1] giving the speed of the Alwac III as compared with that of the *650* as 59% in "normal" mode and 115% in "optimum programming." But the relative storage capacities of these two computers are significant and pertinent in handling the authors' problem and mine. The *650* has a storage capacity of 2,000 words, which can be either instructions or data. Our Alwac III, with 4,000-word drum, can store 8,000 instructions or 4,000 items of data. We actually have in operation on the Alwac III one program consisting of 5,200 instructions, 390 items of data, and 520 program constants. With an 8,000-word drum, programs of over 10,000 instructions are practicable.

There are perhaps two basic reasons for going to a larger than medium machine, say a *705* versus a *650:* faster speed and larger storage capacity.

With our Alwac III there is so much storage capacity that we are not forced to use a larger machine at somewhat more than ten times the cost. Since we can handle all the calculations needed in the foreseeable future on a 1-shift basis, it would not be economical for us to multiply our costs by 10 in order to increase the speed by a factor of 10, or even 50 or a 100.

We point out for the benefit of others with similar problems, who have been prevented from using automatic electronic computing because of the tremendous cost of the large-scale high-speed machines such as an IBM *704* or *705*. Similar storage capacity for needed long programs may be obtained at a small fraction of the cost of these machines.

The argument is often advanced that the low storage capacity of the *650* is entirely adequate, because the program may be worked in parts. The authors have brought out clearly that often this is not the case. To accomplish synthesis design, very long loops are needed. Given a machine that will store, say, 1,500 instructions and 500 words of data, the maximum length of loop is 1,500 instructions, and this frequently is not enough.

The authors mention a program for transformer reactance calculations that contains 2,000 instructions. I should be interested in knowing what lengths of programs the authors are using.

We particularly agree with the authors conclusion that applications should be selected which have a high degree of repetitiousness. The selection of type and size of computer depends upon the applications considered but, as we have pointed out, it is not necessary to go to a million-dollar machine to obtain high storage capacity. In our experience, we too have found that the closed shop appears more advantageous. The over-all contribution and the effect on personnel have been substantially the same as reported by the authors, but the cost of installation is so much less that the economic justification for it is more readily apparent and we have not had to go to such great lengths to justify the economic basis of the installation.

Reference

1. Applied Digital No. 4: Small-Scale Scientific Calculators, **J. W. Carr, III, A. J. Perlis.** *Control Engineering*, New York, N. Y., vol. 3, no. 3, Mar. 1956, pp. 99–104.

Ned Chapin (Stanford Research Institute, Stanford, Calif.): The authors say that there are some computer applications which cannot be done well on small computers; that to handle some applications adequately requires large-scale computational and data-processing capacity. This is also true of some business data processing applications, especially those which attempt to improve the quality of information flowing to management.

The authors suggest a threefold organizational structure for effective computer utilization, and a closed-shop organization of programming effort. These suggestions apply well to many data-processing installations, especially those in which scheduling of data-processing accomplishment is critical.

In the section "Economics" the facts are summarized concerning the introduction of an automatic computer being typically accompanied by a complex set of changes in company procedure. The point is made that because these changes occur as a "package," it is difficult to associate any specific dollar figure to the specific contribution of just the automatic computer.

In making feasibility studies and in justifying the continued use of automatic computers in data-processing installations in business, it has been my observation that management frequently want something that they can point to as indicating that they are "doing the right thing." Often, in order to be secure in the knowledge that the computer installation will do well, management wants to have a "saving" figure to point to, and say, "This is evidence that things are (or will be) in good shape."

Because of the complex set of changes noted previously, the paper suggests talking with management in terms of a "net contribution," not a savings. Can the authors tell us, based on their experiences, the response of their management to this concept? If intangibles are the basis for using the automatic computer, it would relegate a net contribution figure to a position of mere "window dressing." If management has objections to the use of the concept of a net contribution, how would the authors meet this objection?

As to the costs of operating an automatic computer, the literature on costs of programming is referred to, but the authors do not indicate what their cost experience has been. Could we have a summary of what they would include in the cost of programming, and of their experience, on a cost-per-instruction, written-and-debugged basis, for various types of programs?

The discussion of the intangible benefits obtainable through the use of an automatic computer applies equally well to most data-processing installations, and I would agree that these intangible benefits are sufficiently substantial so that they should be kept clearly in mind and emphasized.

P. A. Abetti and S. B. Williams: We wish to thank the discussers for their interesting and thought-provoking discussions. Time and space do not permit us to answer them in detail, so we will limit this closure to the most important points. We are gratified that the data given in the paper verify a posteriori the rule first derived by Dr. Grosch in 1953, a time which belongs to the early history of digital computers. In fact, square roots of the relative speeds given in Table II of the paper are 0.15 for the CPC and 0.016 for the *705*. These values are in good agreement with the relative costs per calculation given also in Table II, namely 0.11 for the CPC and 0.013 for the *705*.

Dr. Grosch discusses also the service bureau approach for various computer applications, which has been used successfully in many installations. In general we agree with his conclusions, but it should be noted that the sharing of the charges may turn out to be unfair, if major programming methods, important subroutines, and utility programs are to be developed first. In this case, the first application on the machine will take the expense and the brunt of this major effort, and successive applications by other groups will require comparatively much less time, effort, and expense. This is why we feel that the economics of applications using common portions of programs should be considered as a whole, since this will justify some applications which would not be justified if considered by themselves. Furthermore, the separate approach to the computer by the various groups may result in an obstacle to integration, and this, as is well known, may be a serious drawback, technically and economically.

Mr. Veinott's discussion is very interesting because it presents experience with a smaller and considerably less expensive computer than the *705* in the field of design of electric machinery. It is clear that there is no need to go to a larger computer, if the one used has sufficient speed and capacity for the applications considered. On the other hand, as is also confirmed by Dr. Grosch, in many cases the load builds up much faster than expected, and many problems, such as reprogramming for conversion, may be avoided by obtaining from the beginning some extra speed and capacity. As mentioned in the paper, a great variety of engineering programs are now being run on the Pittsfield *705*. Some of them, written according to the Print I programming system, may have a few hundred instructions. Typical examples are the determination of the real and complex roots of polynomials, the inversion of matrices, the evaluation of complicated mathematical expressions, curve fitting by the least square method, etc. Others, written in symbolic language or autocoder, may have up to several tens of thousands of instructions.

Mr. Chapin's discussion is also of great interest, since it emphasizes the similarity between engineering and data-processing applications of digital computers, and shows that many problems and solutions are common to both cases. A typical and most important example is the problem of evaluating or measuring the contribution of the computer, for the benefit of management. In our case, the proposed installation of the *705* in the power transformer department was carefully reviewed by management and by the appropriate financial components, according to the established procedures of the General Electric Company. Similarly, the progress obtained since February 1958, when the *705* started operating, is being reviewed constantly according to these procedures. In the process of furnishing the required data, and in many discussions with their associates, the authors felt that it would be advantageous to supplement (but not to replace) this method of evaluation by another method, which would overcome some of the difficulties mentioned in the paper, and of which Mr. Chapin is well aware. For these reasons they proposed the concept of "net contribution," which has been favorably received. It should be emphasized again that, as mentioned in the paper, this is no substitute for the "savings" which every company will want to determine according to its established procedures.

As discussed in the paper, the costs of programming vary over an extended range depending upon the type of application, the computer, the programming system employed, the ability of the programmer, the input and output requirements, the integration between various programs, etc. For this reason we only indicated the range, feeling that specific figures would not be very significant. In the cost of programming we included: 1. the procedural work required to define the problem, the input and output requirements and the general method of approach; 2. numerical analysis work, which may be required for certain engineering problems; 3. preparing rough and detailed flow charts; 4. coding; 5. key-punching and verifying the program and supporting data; 6. conversion to tape of program and supporting data; 7. assembly; 8. debugging; and 9. final verifications of results by trial runs. There is a question whether the cost per instruction is an adequate measure of the cost of programming. Consider, for instance, two equivalent programs, a "tight" one with relatively few instructions, many loops, and no repetition; and another with many instructions, few loops, and several repetitions. Both programs perform the same function and probably will cost about the same, if prepared by programmers of the same ability. The first, however, will have a higher cost per instruction than the second. In a tight program the total cost may be proportional to the square rather than to the first power of the number of instructions, and therefore the cost per instruction will depend upon the length of the program.

For all these reasons, we can answer Mr. Chapin's question only by stating that our experience has yielded an extended range of dollars per instruction. At one extreme are relatively simple mathematical programs written in the Print I system, which cost less than $1.00 per instruction. On the other extreme are complex engineering data-processing programs, which require considerable procedural work, detailed and complex input and output formats, and a considerable amount of supporting data. Some of these were revised, partially or completely, one or more times and their cost amounted to several dollars per instruction, which verifies the values from $4.00 to $10 quoted in the literature. Of course, all intermediate conditions and corresponding values are possible.

In conclusion, we should like to stress that the art of applying digital computers to engineering and data-processing problems is still in its infancy, and that progress can only be furthered by a free and frequent exchange of ideas, data, and experience among the many users of this powerful new tool. For this reason, we are grateful to the discussers for their contribution and hope that out paper will be followed by many others in the broad area it attempted to cover.

Surge Measurement Errors Introduced by Coaxial Cables

J. H. PARK
MEMBER AIEE

NEARLY all surge or impulse voltage measurements involve the use of a cathode-ray oscillograph (CRO) and measuring cable. If an ideal cable is assumed, i.e., one with no losses and a surge impedance which does not vary with frequency, conditions giving zero measurement error can readily be set up. However, all cables have some attenuation and their surge impedance is never the same for all frequencies.[1,2,3] In low-voltage impulse work (where very short cables can be used) such factors may be negligible under certain conditions, but for measuring high-voltage surges a fairly long cable (50 feet or longer) is usually required and possible errors caused by the cable should be considered, especially for measuring steeply rising surges. The purpose of this paper is to discuss such errors and devise methods for measuring and minimizing them.

Attenuation Errors

Cable losses are caused by the resistance of the center conductor and outside sheath together with conductance and dielectric hysteresis in the insulating material between conductors. Losses are usually expressed as an attenuation in decibels per 100 feet at a fixed frequency and in general the logarithm of attenuation in decibels increases in direct proportion to the logarithm of the frequency. For any given surge it is theoretically possible to determine the change in peak value and waveform caused by cable attenuation by using the known values of attenuation at various frequencies and the Fourier integral procedure. However such a procedure is time consuming and would have to be done for each different waveshape of applied surge. Also, it is not completely reliable unless the variation of phase shift with frequency is also taken into account.

An experimental approach is much easier to apply and probably more reliable. One method is to arrange the surge generating circuit so that the pulse to be measured, or one similar to it in shape and rate of rise, can be accurately repeated from shot to shot. (The peak value can be repeated by using a linearly rising voltage chopped with a sphere gap.) Oscillograph records are then taken with (1) a measuring cable of minimum length L, (2) a cable twice this length $2L$, and (3) a cable three times this length $3L$. If the oscillograms obtained under these three conditions are appreciably different in waveshape or peak value, a cable having lower attenuation should be used or a correction for attenuation computed. Polyethylene coaxial cables are used to obtain low attenuation and attenuation can be reduced by using a cable of minimum length and one whose center conductor is large.

For surges which, in general, rise linearly and are chopped by the sparkover of a gap, a correction for the error in peak-voltage measurement due to cable attenuation can be determined. This is done by obtaining values of peak voltage with the three different cable lengths, L, $2L$, and $3L$ as indicated. The decrease in peak voltage ΔV when the cable length is changed from L to $2L$ should be the same as that obtained when the cable length is changed from $2L$ to $3L$, within the accuracy of the measurements. The average of the two values of ΔV is added to the peak voltage obtained with a cable of length L to correct for attenuation.

This method of correcting for cable attenuation was used during an investigation of the change in sparkover value of standard 12.5-cm (centimeter)-diameter spheres set 6 cm apart as the rate of rise of the applied voltage was increased. Values of ΔV were obtained for rates of voltage rise from 10 to 10,000 kv per μsec (microsecond). They were computed as a percentage of total voltage and plotted against rise time (time to sparkover) as obtained by dividing peak voltage by rate of rise. Results are shown in Fig. 1 for *RG-58/U* and *RG-8/U* coaxial cables.

Errors Due to Cable Termination

For most impulse and surge work the CRO cable is terminated by a noninductive resistor which is considered to be equal to its surge impedance in ohms and thus to produce no errors due to reflections. This assumption is not strictly justified, because the surge impedance of all cables varies with frequency and reflections due to different types of surges cannot all be prevented by a terminating resistor of fixed value. The change in impedance with frequency is introduced by (1) the cable resistance and conductance which put a $j\omega$ term in the impedance and can only be neglected over certain frequency ranges and (2) the change in cable inductance caused by nonuniform distribution of current in the sheath and center conductor as frequency is increased. This last factor is actually due to skin effect which also causes a change in resistance with frequency and further complicates factor 1.

Thurston[3] gives a plot in arbitrary units showing the variation of characteristic impedance with frequency for nominal 50-ohm polyethylene cables. Below 100 kc the impedance is given as constant and equal to 55 ohms. From 100 kc to 10 megacycles there is a gradual change from 55 to 50 ohms and it remains at 50 ohms for higher frequencies. Thus if a 50-ohm terminating resistor is used with such a cable (as would normally be

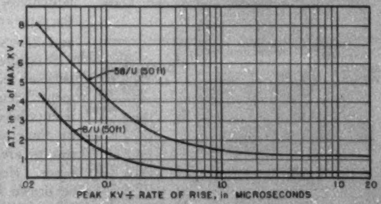


Fig. 1. Attenuation of linearly rising chopped surges by RG-58/U and RG-8/U coaxial cables

---

Paper 58-110, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 9, 1957; made available for printing November 8, 1957.

J. H. PARK is with the U. S. Department of Commerce, National Bureau of Standards, Washington, D. C.

The author acknowledges the assistance of H. N. Cones and J. L. Mills of the National Bureau of Standards who performed the experimental work reported in this paper.

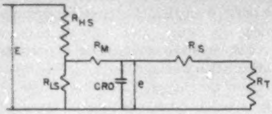


Fig. 2. Circuit used for computing divider ratio with direct voltage applied

the case since such cables are intended for high frequencies and their surge impedance values are given for these high frequencies) there would be a missmatch of 5 ohms for frequencies below 100 kc. This could cause an error of 5% when measuring slowly rising surges, square waves, or 1.5×40-μsec full-wave surges. If a higher value of resistance is used for the termination, say 55 ohms, which would reduce the measurement error for slowly rising surges or full waves, then errors up to 5% would be introduced in measuring steeply rising chopped waves.

In order to measure the actual values of these errors, a better method of terminating the measuring cable at the CRO deflecting system is needed. If another length of the same type of cable is used as a termination, the only discontinuity at the oscillograph would be that caused by the capacitance of the deflecting plates, which is negligible except for very steeply rising surges. An extremely long cable cannot be used because of the effect of its resistance (to be considered shortly), but a sufficient length can probably be used so that reflections of steeply rising chopped surges do not get back to the oscillograph until peak voltage has been reached. Thus a terminating cable with a resistor at its end whose value is chosen to give no reflection for slowly rising surges should serve to eliminate termination errors for all types of surges. The effect of the resistance of a terminating cable will first be considered, then an experimental method for deciding when such a cable is needed and its length will be described.

Table I. Change in Divider Ratio with Time Introduced by Resistance of Terminating Cable for a Suddenly Applied Voltage

Computed by Applying Equation 3 to the Case of a Very-Long RG-58/U Terminating Cable with R_{HS}=1,000 Ohms and R_{LS}= 10 Ohms

Time After Application of Voltage, μsec	Divider Ratio $e_a(t)/e(o,t)$
0	119.5
0.2	119.2
0.4	118.9
0.6	118.6
0.8	118.2
1.0	118.0
1.5	117.4

The effect of the resistance of a terminating cable on CRO measurements depends upon the circuit to which the CRO measuring cable is connected. A typical measuring circuit is shown in Fig. 2 where resistance only is considered since a d-c applied voltage is assumed. The low side of a resistance divider ($R_{HS}+R_{LS}$) is shown connected to the deflection plates of a CRO by a measuring cable of resistance, R_M. The measuring cable is terminated by a longer cable of resistance, R_S, which in turn is terminated by R_T.

For a d-c applied voltage with $R_M=0$:

$$e=E\frac{R_{LS}}{R_{LS}+R_{HS}\left(1+\dfrac{R_{LS}}{R_T+R_S}\right)} \tag{1}$$

The error introduced by neglecting R_S is small if $R_S<<R_T$ and it is considerably reduced if $R_{LS}<<R_T$. This d-c computation gives the maximum error likely to occur due to resistance of the terminating cable during surge measurements. An actual computation is possible for a surge voltage applied at E (Fig. 2) if the general solution of transmission-line equations (which take cable resistance into account) is used. Weber[4] has indicated, in general, how to apply LaPlace transform methods to such problems and he has shown that for a suddenly applied voltage, the current taken by an infinitely long cable decreases with time, i.e., in effect the surge impedance increases with time. By using similar methods, a solution was derived and is given in Appendix I for the voltage $e(o,t)$ at the deflecting plates of a CRO which are connected by a negligibly short measuring cable to the low side of a resistance divider, when voltage is suddenly applied to the divider high side. The measuring cable is assumed to be terminated by an infinitely long length of cable of the same type and its resistance, inductance, and capacitance are considered. The result is derived as an infinite series and five terms, which give sufficient accuracy if $R_{LS}\leqq 10$ ohms, are given by equation 3. If cable resistance is negligible ($r=0$) the divider ratio does not change with time and equation 3 gives the same results as equation 1 with $R_S=0$ and $R_T=\sqrt{l/c}$.

It is difficult to deduce general relations between cable resistance and measurement accuracy from these theoretical results because of the assumptions made in their derivations and because the results are in the form of an infinite series. However, some conclusions can be drawn by applying equation 3 to a specific case. Computations were made for *RG-58/U* coaxial cable, $R_{HS}=1{,}000$ ohms, and $R_{LS}=10$ ohms. Results are shown in Table I. At zero time the divider ratio is the same as that obtained on direct current with the cable replaced by a resistor equal to its surge impedance. Divider ratio decreases as time increases and at infinite time reaches the value that would be obtained on direct current with an open circuit at the CRO. For a finite length of cable it can be assumed that the voltage at the CRO would be the same as for an infinite cable up to the time when the first reflection from the end of the cable gets back to the CRO. For a terminating resistor at the end of this cable equal to its impedance as obtained by neglecting resistance, this reflection would be such as to eliminate any further change in divider ratio. Thus for a 500-foot length of *RG-58/U* cable the maximum change in ratio for a suddenly applied voltage would be that given in Table I for time equal twice the pulse travel time or 1.5 μsec. This corresponds to an error of 1.8%, which agrees with the value obtained from d-c considerations using equation 1 and saying the error is the difference found in computing e for $R_S=0$ and $R_S=7$ ohms (the resistance of 500 feet of *RG-58/U*).

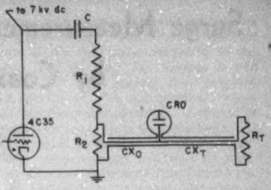


Fig. 3. Circuit used in measuring effect of terminating cable on divider ratio

In order to obtain an experimental verification of the error caused by cable resistance, the setup shown in Fig. 3 was used. Capacitor C was charged to about 7 kv and then discharged through noninductive resistors, R_1 and R_2, by the hydrogen thyraton. The voltage across R_2 was measured by connecting a 45-foot length of *RG-58/U* coaxial cable, CX_0, from R_2 to the CRO deflection plates. The termination for cable CX_0 consisted of either a 500-foot length of *RG-58/U* coaxial cable, CX_T with R_T at its end, or a 3-foot length of *RG-58/U* coaxial cable with R_T at its end. R_T was a noninductive resistance of 55 ohms which equals the surge impedance of the cable for slowly rising surges or full waves. In order to obtain a direct comparison of

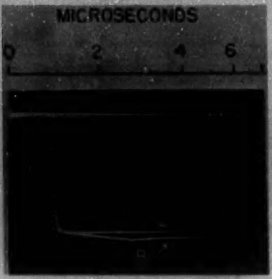


Fig. 4. Oscillograms showing effect of resistance of terminating cable (RG-58/U) obtained with setup shown in Fig. 3 with $R_1=150$ ohms and $R_2=50$ ohms

A—Obtained with $CX_T=3$ feet
B—Obtained with $CX_T=500$ feet

voltage for the two lengths of terminating cable, repeat shots were taken on the same oscillogram with everything the same except terminating cable length. An example is shown in Fig. 4, obtained with $R_1=150$ ohms and $R_2=50$ ohms, which gives a difference large enough to be readily noticed. The long terminating cable gives a higher voltage at the CRO as seen by comparing the two traces. The maximum difference is about 5% as determined from measurements on the oscillogram. By using equation 1 which was derived for a steady d-c applied voltage and the measured value of cable resistance (7 ohms), the change in measured voltage introduced by cable resistance can be computed and it agrees with the experimentally determined value to within 0.5%.

Experimental Check of Cable Termination Errors

Measurement errors introduced by the method used to terminate a CRO measuring cable depend upon the type of surge being measured and the values of resistance used for the divider. They cannot be assumed to be negligible. While theoretical computations are useful in estimating such errors the final check on their magnitude must be obtained experimentally.

For measuring 1.5×40-μsec full waves or those chopped on the tail, the cable surge impedance may be considered as the value obtained at low frequency (below 100 kc). Thus a fixed resistor located at the CRO can be used to terminate the cable. A circuit similar to that shown in Fig. 3 should be used to determine the exact value of R_T. R_1 and R_2 should be noninductive resistors and the value of R_2 should be approximately equal to the surge impedance of the measuring cable to prevent confusion caused by reflections from the input end of the cable. The terminating cable CX_T should be at least 100 feet long and several different values of R_T should be tried so that the one giving minimum reflections can be chosen. This was done for *RG-58/U* and *RG-8/U* cables; the results are shown by the oscillograms of Figs. 5 and 6. The reflections are easily identified as small pips on the flat top of the suddenly applied voltage. If the pip introduces a small step discontinuity which persists for several microseconds, it indicates that R_T is too large if the step increases the measured voltage, or too small if it decreases the voltage. When R_T is exactly right there will be no step discontinuity but usually a pip of short duration which is a reflection of the high-frequency components associated with the sudden change in voltage. From the measured magnitude of the step discontinuity a value for the correct terminating resistor (for full-wave measurements) was obtained from each CRO record. For each type of cable, measurements were obtained on ten or more oscillograms and the maximum variations from the average value were ±1 ohm. The average values were 55.7 ohms for *RG-58/U* cable and 52.7 ohms for *RG-8U* cable.

For measuring steeply rising surges, chopped on their rising front, a terminating resistor whose value had been determined as described would introduce an error due to reflection of the high-frequency components of surge. Thurston's[3] data on 50-ohm cables would indicate that this error might be as high as 5%; however, it would be expected to vary considerably with rate of voltage rise or time to chopping and probably to some extent with cable type and construction. This error was determined experimentally for *RG-58/U* and *RG-8/U* cable by obtaining CRO records of linear rising chopped surges of various durations using both a 3-foot and 200-foot terminating cable. The resistor, R_T, at the other end of the cable was made approximately equal to the full-wave surge impedance of the cable as given in the preceding paragraph. For each rate of voltage rise, peak voltage was duplicated on successive shots by using 12.5-cm spheres spaced 6 cm for chopping. For the longer duration surges the measured peak value was the same for both 3- and 200-foot terminating cables. As the duration of the surge was decreased, a point was reached where higher measured peak values were obtained with the 3-foot cable than with the 200-foot cable. This

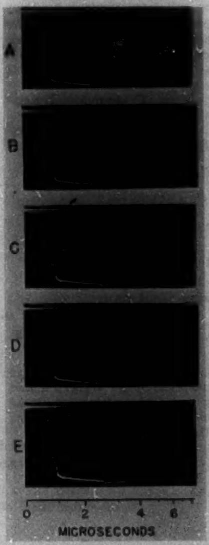


Fig. 5 (left). Oscillograms showing reflections from the end of a 500-foot-long RG-58/U cable, terminated with various values of R_T, as given in ohms below. Obtained using setup shown in Fig. 3 with $R_1=150$ ohms and $R_2=50$ ohms

A—51.4
B—53.6
C—54.6
D—55.4
E—57.9

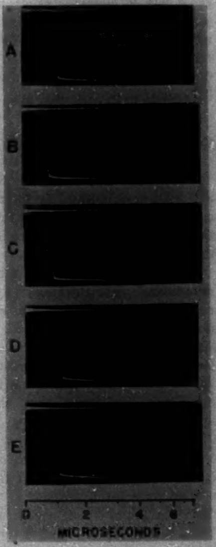


Fig. 6 (right). Oscillograms, as described in Fig. 5, for RG-8/U cable

A—49.8
B—51.4
C—52.4
D—53.6
E—55.4

began to occur for durations of about 0.2 μsec and for 0.1-μsec rise time the difference was 1%. For a time to chopping of 0.05 μsec, differences in measured voltage for the 3- and 200-foot terminating cables were determined for several different values of terminating resistor, R_T. From these results the correct value of R_T for a surge of 0.05-μsec rise time was found to be 53.8 ohms for *RG-58/U* cable and 50.9 ohms for *RG-8/U* cable. Thus the change in surge impedance is only about 2 ohms in going from full-wave to very-steep-wave (0.05-μsec duration) measurements, which is considerably less than the 5-ohm change deduced from Thurston's data. This 5-ohm change in impedance might be attained if shorter duration surges were used.

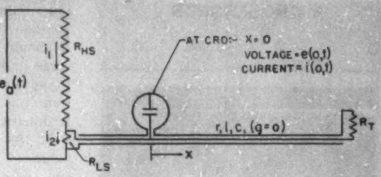
For measuring very short duration surges it would be advisable to use a terminating cable with a wave travel time down and back of 0.2 μsec, i.e., about 100 feet for *RG-8/U* or *RG-58/U* cable. There remains a possible error due to resistance of the terminating cable. Equation 1 can be used to compute an upper limit for this error when the low-side voltage of a resistance divider is being measured. For 100 feet of *RG-8/U* (resistance = 0.3 ohm) this error is less than 0.5% for all values of R_{LS} and R_{HS}. For 100 feet of *RG-58/U* (resistance = 1.4 ohms) this error is somewhat larger, but it would be less than 1% for $R_{LS} \leqq 10$ ohms. Thus a 100-foot terminating cable with R_T equal to its surge impedance for full waves would be suitable for measuring all types of surges. The 100-foot terminating cable is not needed for measuring surges with rise times greater than 0.2 μsec.

Conclusions

The possibility of errors caused by attenuation introduced by the measuring cable and the method used for terminating this cable should be considered in measuring high-voltage surges.

Attenuation errors are quite easily determined experimentally for linear-rising chopped surges and a procedure for doing this has been described. The results obtained indicate that *RG-8/U* coaxial cable has less attenuation than *RG-58/U*. If *RG-58/U* cable is used a correction for attenuation should be applied, even for full-wave measurements.

Fig. 7. Circuit used for computing divider ratio when a surge voltage, $e_a(t)$, is applied



Since the surge impedance of coaxial cables is not independent of frequency, a fixed terminating resistor at the CRO deflection plates is not satisfactory for all types of surges. By using a terminating cable of the same type as the measuring cable, a method was devised to obtain a termination satisfactory for all types of surges. This method was used to measure reflection errors introduced by a fixed terminating resistor at the CRO deflecting plates. It was found that such errors were less than 1% when using *RG-58/U* or *RG-8/U* cable to measure surges with rise times (times to chopping) greater than 0.2 μsec.

Appendix I. Derivation of Divider Ratio for a Suddenly Applied Voltage with Terminating Cable Constants r, l, and c Considered

Referring to the notation given in the circuit diagram of Fig. 7, the purpose of this derivation is to obtain an expression for $e(o, t)$ in terms of $e_a(t)$ and the circuit constants. The measuring cable, from R_{LS} to CRO, and the terminating cable, from CRO to R_T, are assumed to be taken from the same reel. The measuring cable is short enough so that its resistance can be neglected. The terminating cable is very long and terminated by R_T, which is equal to its surge impedance.

Referring to Fig. 7, the following equation can be written:

$$i_1 = i_2 + i(o, t)$$

$$e_a(t) = [i_2 + i(o, t)]R_{HS} + i_2 R_{LS}$$

$$i_2 R_{LS} = e(o, t)$$

From these equations

$$e_a(t) = e(o, t)\left[1 + \frac{R_{HS}}{R_{LS}}\right] + i(o, t)R_{HS}$$

Its Laplacian transform is

$$Le_a(t) = E(o, p)\left[1 + \frac{R_{HS}}{R_{LS}}\right] + I(o, p)R_{HS} \quad (2)$$

$E(o, p)$ and $I(o, p)$ are given by equations 17 and 18, respectively, of Weber's article on traveling waves on transmission lines.[4] For an infinite line with no initial voltage or current, these equations reduce to

$$E(o, p) = B$$

and

$$I(o, p) = \frac{B}{Z_c}$$

where B is a constant and

$$Z_c = \sqrt{\frac{r + lp}{g + cp}}$$

Substituting these in equation 2 and solving for B gives

$$B = Le_a(t)\frac{1}{1 + \frac{R_{HS}}{R_{LS}} + \frac{R_{HS}}{Z_c}}$$

Let

$$1 + \frac{R_{HS}}{R_{LS}} = N$$

then

$$E(o, p) = Le_a(t)\,\frac{Z_c}{NZ_c + R_{HS}}$$

For a suddenly applied voltage

$$Le_a(t) = \frac{1}{p}$$

and for

$$g = o,\ Z_c = \sqrt{\frac{lp + r}{cp}}$$

$$\mathrm{E}(o, p) = \frac{1}{p}\,\frac{\sqrt{lp + r}}{N\sqrt{lp + r} + R_{HS}\sqrt{cp}}$$

To find the inverse transform of $E(o, p)$ which is the voltage at the CRO, use the binomial theorem to expand the fraction in the foregoing equation giving

$$E(o, p) = \frac{1}{pN} - \frac{R_{HS}}{pN^2}\left[\frac{cp}{lp + r}\right]^{1/2} + \frac{R_{HS}^2}{pN^3}\left[\frac{cp}{lp + r}\right] - \frac{R_{HS}^3}{pN^4}\left[\frac{cp}{lp + r}\right]^{3/2} + \ldots$$

Take the inverse transform of each term on the right side of this equation to obtain $e(o, t)$. Going as far as available tables of transforms allow gives

$$e(o, t) = \frac{1}{N} - \frac{R_{HS}}{N^2}\left[\frac{c}{l}\right]^{1/2} \epsilon^{\frac{-rt}{2l}} I_0\left[\frac{rt}{2l}\right] + \frac{R_{HS}^2}{N^3}\,\frac{c}{l}\,\epsilon^{\frac{-rt}{l}} - \frac{R_{HS}^3}{N^4}\left[\frac{c}{l}\right]^{3/2} \epsilon^{\frac{-rt}{2l}} \times \left\{-\frac{rt}{l} I_1\left[\frac{-rt}{2l}\right] + \left[1 - \frac{rt}{l}\right] I_0\left[-\frac{rt}{2l}\right]\right\} + \frac{R_{HS}^4}{N^4}\,\frac{c^2}{l^2}\left(1 - \frac{rt}{l}\right)\epsilon^{\frac{-rt}{l}} + \quad (3)$$

In this equation I_0 and I_1 are hyperbolic Bessel functions.

References

1. Transient Analysis of Coaxial Cables Considering Skin Effect, R. L. Wigington, N. S. Nahman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, Feb. 1957, p. 166.

2. The Development of a Precision Termination for 0.375-Inch Polythene-Disc-Insulated Coaxial Cable, R. J. Cheetham, E. L. Mather, W. W. H. Clarke. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101-IV, 1954, p. 135.

3. The Measurement of Cable Characteristics, W. R. Thurston. *General Radio Experimenter*, Cambridge, Mass., May 1957.

4. Traveling Waves on Transmission Lines, Ernst Weber. *Electrical Engineering*, vol. 61, June 1942, pp. 302–09.

◆

Discussion

Charles J. Miller, Jr., and **John F. Wittibschlager** (The Ohio Brass Company, Barberton, Ohio): The author presents a very concise and clear analysis of the two types of errors which may be introduced by the use of a coaxial cable to connect the potential divider to the cathode-ray oscillograph.

The attenuation characteristics of a 100-foot 3-shield coaxial cable, with electrical characteristics similar to those of an *RG-11/U* 75-ohm cable were run in the Barberton High Voltage Laboratory of the Ohio Brass Company[1] using a technique very similar to that described by the author.

The results of this investigation are compared with those of the author in Fig. 8. This comparison indicates a close resemblance between the 3-shield cable tested and an *RG-8/U* cable. The somewhat higher attenuation of the *RG-11/U* over that of the *RG-8/U* at longer times to flashover is explained by its somewhat higher d-c resistance resulting from its somewhat smaller central conductor. At higher frequencies, on the other hand, the *RG-8/U* cable has approximately a 10% greater attenuation in decibels per 100 feet than does the *RG-11/U* cable, according to published catalogue data. This difference in attenuation would account for the *RG-11/U* curve crossing over the *RG-8/U* curve at shorter times to flashover.

Since the maximum deviation between the curves for the *RG-8/U* and *RG-11/U* cables represents a difference in attenuation of 0.6 of 1%, which is apparently explained by the geometric and electrical differences in the two cables, it would appear that the results obtained in the two different laboratories agree very well.

In the Barberton tests, the cable terminating resistance at the CRO plates was carefully matched for a time to flashover of about 0.08 μsec, using linearly rising chopped surges which were produced by the flashover of a 6-cm gap between 12.5-cm spheres. The principal elements of the arrangement are shown in Fig. 9. As indicated, the divider consists of a 1,000-ohm noninductive resistor R_1 and a 1-ohm noninductive resistor R_2, the details of which are described completely in a paper by the discussers.[1] The 3-shield coaxial cable is terminated by a noninductive resistor R_3 located at the CRO plates.

Resistor R_3 is made of a suitable length of Karma ribbon compressed in a zigzag fashion with two layers of 0.001-inch Mylar tape between adjacent layers. Both the coaxial cable and the terminating resistor R_3 are connected directly to the CRO plates with leads less than 1 inch long. A number of similar R_3 resistor units were made, each with a slightly different value of resistance about the 75-ohm level.

The results of the tests are shown by the oscillograms in Fig. 10. Note that only the lower voltage portion of each oscillogram has been reproduced. For the value of R_3 that terminates the coaxial cable properly for the 0.08-μsec pulse no reflected wave is set up, and hence there can be no voltage error due to improper matching. However, for other values of R_3, there will be a small reflected wave, the magnitude of which will depend upon the degree of mismatch, and thus a voltage error may exist at the CRO plates. This first reflected wave will travel to the input end of the cable where it encounters a very low impedance of about 1 ohm. A nearly equal wave of opposite polarity will be generated due to reflection, and this second wave will return to the CRO plates in about 0.3 μsec, the round-trip travel time in a 100-foot cable. Being of opposite polarity and nearly equal magnitude to the first reflected wave, it will nearly cancel the voltage error caused by the initial reflection. The return of the wave of opposite polarity will be detected by a sudden small change in the deflection of the oscillogram at about 0.3 μsec.

In Figs. 10(A) and 10(C), the voltage pips are due to a mismatch between the terminating resistance R_3 and the cable surge impedance, while Fig. 10(B) shows no such pip since the terminating resistance of 74.7 ohms provides the proper value for matching the cable surge impedance for the applied 0.08-μsec pulse.

As a matter of interest, the pips in Figs. 10(D) are due to the use of a commercially available noninductive resistor which has approximately the same d-c resistance as that used to obtain Fig. 10(B). The pips are probably due to mismatch caused by the residual inductance in this resistor.

The author points out that if the coaxial cable is terminated in that value of resistance which provides proper matching for a $1^1/_2 \times 40$-μsec wave, it will be somewhat in error for waves that crest in times shorter than 0.2 μsec, due to a slight reduction in cable surge impedance for these shorter waves, and hence a mismatch of terminating resistance. To avoid this difficulty, he proposes the use of a terminating impedance consisting of a 200-foot extension of the coaxial cable past the CRO plates, with a terminating resistance which matches properly for long times to flashover. However, the series resistance of the 200-foot cable extension introduces additional complications, as pointed out by the author.

It is believed that using a terminating resistor at the CRO plates which matches the surge impedance of the cable for short times to flashover will not introduce appreciable error for waves which reach crest values at times in excess of the round-trip travel time in the cable itself, provided the low-voltage resistance R_2 of the potential is small compared to the surge impedance of the cable.

As discussed in the foregoing, the return of an opposite polarity wave at the round-trip travel time in the cable nearly wipes out the voltage error created by the mismatch between cable surge impedance and terminating resistor R_3. Any residual error

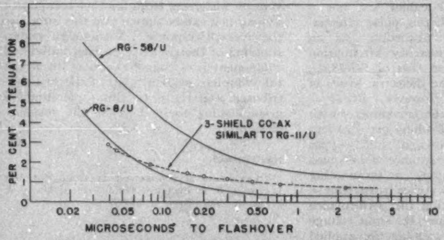


Fig. 8. Comparison of attenuation characteristics of 50-foot lengths of coaxial cables on linearly rising chopped surges

Curves for RG-58/U and RG-8/U from Fig. 1 of the paper
Curve for 3-shield coaxial cable from test at Ohio Brass High Voltage Laboratory

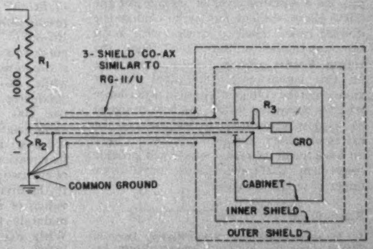


Fig. 9. Test arrangement used to verify match between surge impedance of coaxial cable and terminating resistor R_3

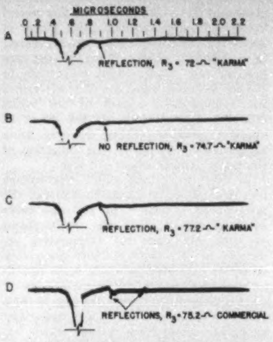


Fig. 10. Oscillograms showing effect of various terminating resistors at CRO plates

that may exist will certainly be reduced still more after a second round in the cable. By the time a $1^1/_2 \times 40$-µsec wave reaches its crest value, the error due to mismatch must be negligibly small.

This leaves the possibility of error for pulses which reach their crest value between the time for which the terminating resistor R_3 is properly matched and the round-trip travel time in the cable, which in the example given is from 0.08 to 0.30 µsec. However, when we remember that for 0.08-µsec pulses the cable is properly terminated, and as the author points out the change of surge impedance from full wave to 0.05-µsec wave is only about 2 ohms out of 50 ohms, or about 4%, the magnitude of the mismatch at 0.30 cannot be very large, and hence the errors in this range must be quite small.

There is also the possibility of error due to mismatch for pulses which reach their crest value at times shorter than that for which the cable is properly terminated. Here again, the error may be quite small since the surge impedance of the cable probably does not change too much for pulses between 0.05 and 0.08 µsec.

The terminating resistance at the CRO plates offers a practical advantage that is not available with the extended cable method of termination, and that is the possibility of a potential tap in order to provide a suitable level of voltage for the CRO plates without having to change the divider ratio. Such a resistor with a 0.4 tap was made in Barberton, using the zigzag Karma ribbon resistor, and tests on this potentiometer indicated that it gave accurate voltage division in accordance with its measured ratio. As it is designed as a plug-in unit, lead lengths can be extremely short, and it is possible to have an assortment of ratios which will provide convenient CRO deflections for a wide range of input voltages.

Reference

1. Measurements of Steep-Front Impulse Waves with an Isolated Screen Room Installation, **C. J. Miller, Jr., J. F. Wittibschlager.** *AIEE Transactions*, vol. 77, pt. I, July 1958, pp. 262–71.

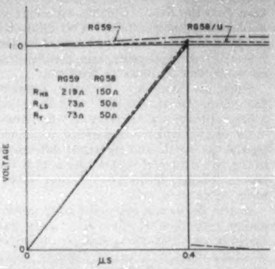


Fig. 11. Comparison of cable input voltages for 130-foot lengths of RG-59/U and RG-58/U terminating cables. Step function and steep-front waves applied

F. A. Fisher and **J. S. Kresge** (General Electric Company, Pittsfield, Mass.): Generally, when the transmission-line equations are solved, the effects of series resistance, shunt conductance, and skin effect are neglected. The resulting solutions then imply that the transmission line has a surge impedance independent of time and frequency, that there is no attenuation, and that a terminating resistance equal in magnitude to the surge impedance of the cable will prevent all reflections. This paper shows that none of these conclusions are strictly correct. Mr. Park discusses attenuation, waveshape distortion, determination of surge impedance, and terminating methods. We should like to bring up a few points on these same subjects.

Fig. 1 clearly shows the advantage of using *RG-8/U* cable instead of the smaller diameter *RG-58/U*. Practically without exception one can say that the small-diameter coaxial cables are unsatisfactory for pulse measurement use, unless they are used only in very short lengths. We have obtained similar data for *RG-11/U* and *RG-59/U*, which are the equivalent 75-ohm cables, and this also indicates that the small-diameter cable, *RG-59/U*, is not generally satisfactory.

It is important to realize that the radio-frequency attenuation figures normally published for coaxial cables are not a reliable guide for predicting pulse attenuation and waveshape distortion. As an example, the radio-frequency attenuation of *RG-59/U* is less than that of *RG-58/U*, presumably because the dielectric losses in *RG-59/U* are less. However, *RG-59/U* employs a copperweld center conductor with a d-c resistance of 5.5 ohms per 100 feet while *RG-58/U* employs a solid copper conductor with a d-c resistance of 1.4 ohms per 100 feet. Assuming a terminating cable length of 130 feet (twice travel time = 0.4 µsec) and calculating the performance by equation 1 of the paper, the input voltage of the two cables for a step function applied to the divider is as shown in Fig. 11. The values of R_{hs}, R_{ls}, and R_t were chosen to maintain the same over-all divider ratio. With *RG-59/U* cable the input voltage is 4% too high at 0.4 µsec; with *RG-58/U* cable it is 1.6% too high.

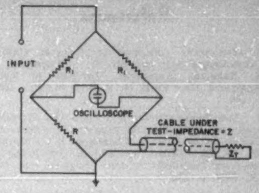


Fig. 12. Impulse bridge circuit for determining correct cable termination

Fig. 11 also shows the corresponding response to a linearly rising wave chopped abruptly to zero at crest time (steep-front wave) as calculated by the methods outlined in another paper.[1] Such a wave cresting at 0.4 µsec or less would be reproduced with an error of half that of the step function response at that time.

When determining the surge impedance of a transmission line or a cable the sensitivity may be improved by using a bridge circuit. Fig. 12 shows a diagram of a simple bridge which we have used. The resistors R_1 are taken approximately equal to the expected surge impedance. If R is adjusted for a balance at $t=0$ then the cable input impedance is equal to R. The correct terminating impedance is found by varying Z_t until no reflections at $t=$ twice travel time are seen. If such a bridge is constructed care should be taken to avoid errors caused by residual inductances and skin effect in the bridge elements.

Mr. Park uses a combination of a terminating cable and a fixed resistor to provide a termination correct for all frequencies. This can be done in other ways also; one simple way which we have found to be quite effective is shown in Fig. 13. R and C may be adjusted experimentally until reflections are minimized.

Finally, it should be noted that an incorrectly terminated signal cable may compensate for errors introduced by some other component of the measurement system. In a complete impulse-voltage-measuring system we are primarily interested in the over-all response and it would be legitimate to introduce errors deliberately in one component that compensate for errors from another component as long as it can be shown that this improves the over-all response. A thorough understanding of the response of each individual component is of course necessary to adjust the complete system. Mr. Park has contributed materially to this understanding in publishing these data on the response of cables.

Reference

1. Transient Response of Impulse Voltage Dividers, **F. A. Fisher.** *AIEE Transactions*, vol. 77, pt. I, 1958 (*Paper no. 58-37*).

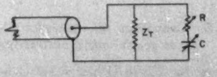


Fig. 13. Frequency-sensitive cable termination using adjustable trimmer elements

A. F. Rohlfs (General Electric Company, Pittsfield, Mass.): The influence that measuring cables may have on the accuracy of voltage divider systems has received much less attention over the years than have the other parts of divider systems. Mr. Park's paper on this subject should be most welcome to those people using or studying divider systems.

In the appendix of the paper a rigorous mathematical solution to the problem is developed. The resulting expression, however, is quite complicated. By using a traveling-wave approach with same approximations a simple solution can be found.

Assume that the cable is made up of n small elements each having a lumped series resistance r as shown at the top of Fig. 14.

The transmission coefficient for each element will be

$$A=\frac{2Z}{r+2Z} \tag{4}$$

and the reflection coefficient will be

$$B=\frac{r}{r+2Z} \tag{5}$$

The reflection coefficient at the input end of the cable will be

$$C=\frac{R_{LS}-Z}{R_{LS}+Z} \tag{6}$$

where R_{LS} is the terminating resistor at the input end. It is assumed that R_{LS} will be small compared to R_{HS}, the resistance of the high-voltage arm of the divider. The output end of the cable will be terminated with its surge impedance, Z.

A lattice diagram for a cable having a travel time T is shown in Fig. 14. The output voltage will be the sum of the components arriving at the right up to the time t under consideration. Thus

$$V_{\text{out}}=A^{(n-1)}+A^{(n-1)}[BC+(n-1)B^2]+ A^{(n+1)}[BC+(n-1)B^2]+\ldots+ A^{\{n-1+2(m-1)\}}[BC+(n-1)B^2]$$

$$V_{\text{out}}=A^{(n-1)}[1+[BC+(n-1)B^2]\times [1+A^2+A^4\ldots A^{2(m-1)}]]$$

$$V_{\text{out}}=A^{(n-1)}\left\{1+[BC+(n-1)B^2]\times \left[\frac{1-A^{2m}}{1-A^2}\right]\right\} \tag{7}$$

where

$$m=\frac{t-T}{2T}n \tag{8}$$

and

$$T\leq t\leq 3T \tag{9}$$

Now if m and n are of the order of 100 or less and $(1-A)$ is 0.001 or less,

$$A^n\cong 1-n(1-A)$$

and

$$A^{2m}=1-2m(1-A) \tag{10}$$

This will be the case for most practical cables in use with high-voltage dividers.

Substituting equation 10 in equation 7,

$$V_{\text{out}}=\{1-(n-1)(1-A)\}\times \{1+[BC+(n-1)B^2]m\} \tag{11}$$

From equatious 4 and 5

$$B=1-A \tag{12}$$

thus

$$V_{\text{out}}=\{1-(n-1)B\}\times \{1+[mBC+m(n-1)B^2]\} \tag{13}$$

For the conditions imposed in the foregoing on m, n, and $(1-A)$, the B^2 term will be small and can be neglected without appreciable error. If the lattice diagram were complete, it would also contain higher powers of B. However, these would be still smaller than the B^2 term and thus may be neglected. Therefore

$$V_{\text{out}}=[1-(n-1)B][1+mBC] \tag{14}$$

This can also be written

$$V_{\text{out}}=A(n-1)[1+mBC] \tag{14A}$$

The input voltage will be the sum of the components at the input end of the lattice diagram up to the time (t) under consideration.

$$V_{\text{in}}=1+B(1+C)(1+A^2+A^4\ldots+A^{2(k-1)})$$

Simplifying as before

$$V_{\text{in}}=1+KB(1+C) \tag{15}$$

where

$$K=\frac{t}{2T}n \tag{16}$$

and

$$0\leq t\leq 2T \tag{17}$$

Because of the approximations used, it is not possible to convert equation 15 into Park's equation 1 for the d-c solution. However, numerical substitution for several conditions to be discussed in the following shows that the equation 15 gives approximately the same answers as does Park's equation.

To illustrate the use of these equations let us assume that the resistance, R, of the cable is 11 ohms, $n=100$, $Z=55$ ohms, and $R_{LS}=5$ ohms, 55 ohms, and 1,000 ohms. Then $B=0.001$ and the input and output waves for the three values of R_{LS} will be shown in Fig. 15. A resistance of 11 ohms represents the resistance of 785 feet of *RG-58/V* cable.

Examination of these waves shows that the shape of the output wave will be true only when R_{LS} is equal to Z; however, the amplitude will be low. The shape of the wave at the input end will approach being correct when R_{LS} is small compared to Z.

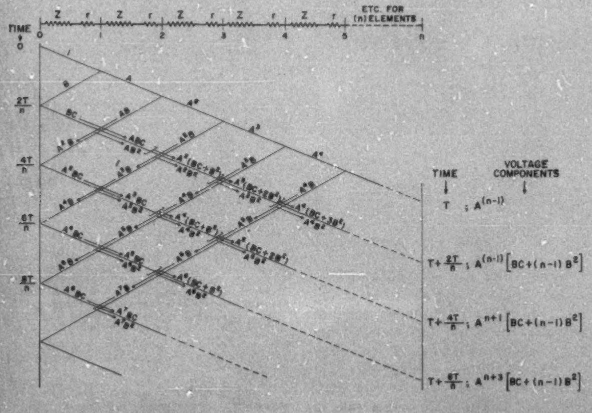
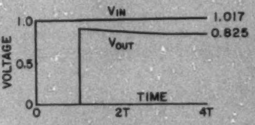
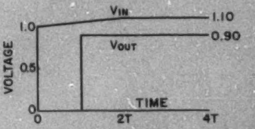

Z
r
ETC. FOR (n) ELEMENTS
TIME
0
1
2
3
4
5
n
2T/n
4T/n
6T/n
8T/n
TIME
VOLTAGE COMPONENTS
T ; A^(n-1)
T+2T/n ; A^(n-1) [BC+(n-1)B²]
T+4T/n ; A^(n+1) [BC+(n-1)B²]
T+6T/n ; A^(n+3) [BC+(n-1)B²]

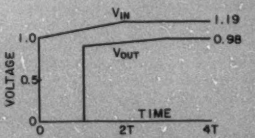
Fig. 14. Lattice diagram of cable subdivided into (n) elements



(a) $R_{LS}=5$



(b) $R_{LS}=55=Z$



(c) $R_{LS}=1000$

Fig. 15. Voltage waves at input and output ends of cable for $R_{LS}=5$, 55, and 1,000 ohms

Table II. Input Voltage With 11 Ohms Resistance in Cable

R_{LS}	Input Voltage to Cable: Equation 15	Input Voltage to Cable: Equation 1 by Park
5	1.017	1.015
55	1.100	1.090
1,000	1.190	1.183

To compare equation 15 to Park's equation 1, it is necessary to have a complete divider. Therefore, let us assume that the divider ratio would be 1,000 to 1 if the resistance of the cable were zero. Then by equation 15 the input voltage with 11 ohms resistance in the cable will be as shown in Table II for 1,000 volts applied to the top end of the divider. Using values of R_{HS} that will give ratios of 1,000 to 1 and substituting in Park's equation, input voltage were also calculated.

The two methods agree fairly well.

Fig. 15 also indicates that the apparent attenuation of a measuring cable is dependent upon the value of R_{LS}. For example, in Fig. 15 the apparent attenuation is 2% for $R_{LS}=1{,}000$, 10% for $R_{LS}=55$, and 18% for $R_{LS}=5$. Thus one must take care that in applying a correction for "attenuation" the correction factor is determined for the actual values of R_{LS} in use.

This approach to the distortion of voltage waves by a cable does not consider the effect of frequency. In so far as higher frequencies cause an increase in cable resistance by means of a skin effect, the equation may be valid provided the proper resistance values are used. However, some of the increased "attenuation" at the higher frequencies may be due to dielectric loss and it is doubtful that such an effect could be handled in the same manner.

J. H. Park: I wish to thank the discussers for their contributions since they all enhance the value of the paper. While surge-voltage-measurement errors, introduced by the measuring cable, are normally quite low and nearly negligible for most laboratory tests, there appears to be considerable interest in methods for measuring such errors and correcting for them.

The method described by Mr. Miller and Mr. Wittibschlager for checking their cable termination resistor by using a steeply rising chopped wave is certainly satisfactory for their particular conditions. However, when the divider low side is much higher than 1 ohm (up to 50 ohms under some conditions) there will be no reflected wave from the low side to compensate for full-wave termination missmatch. A 100-foot terminating cable with full-wave matching resistor at its end, as described in the paper, can be used for all types of surges and all values of divider low-side resistance. By using *RG-8/U* cable the resistance of the 100-foot length used for termination is kept entirely negligible. Such a termination does not permit locating a resistance network at the CRO plates for adjusting sensitivity as used by Miller and Wittibschlager but experience at the National Bureau of Standards has indicated that such a network increases the likelihood of measurement errors due to stray magnetic fields from ground currents in the cable sheath. The magnitude of the deflection at the CRO can rather easily be controlled by using divider low-side resistors of different value.

The bridge circuit suggested by Mr. Fisher and Mr. Kresge (Fig. 12) to determine cable surge impedance would improve sensitivity if either (1) an amplifier is used in the CRO deflecting circuit or (2) an increased voltage is impressed on the cable. In either case, for steeply rising surges, it would appear that special precautions must be applied to eliminate pickup in the deflecting circuit because this circuit is not connected directly to ground. The suggestion by these discussers that errors due to cable termination may compensate for other errors would possibly work out under certain conditions. However, most errors are introduced by variables whose relation to other factors is not definitely known so that two sets of errors which compensate, say, for low-voltage impulse tests may not retain that relation under high-voltage conditions. The best procedure is to isolate errors wherever possible and apply corrective measures for each one individually.

The method presented by Mr. Rohlfs of taking cable resistance into account by using the traveling-wave lattice diagram is very interesting and contributes much to the paper. As indicated by him it is difficult, if not impossible, to work out a computation method which takes all factors, such as skin effect and dielectric losses, into account, and so an experimental check is always desirable. His computations show that cable attenuation as caused by its series resistance depends upon the particular values of resistance used in the divider. For this reason it is advisable to use as short a measuring cable as possible and one with a fairly large center conductor. This also emphasizes the importance of checking cable attenuation by the method described in the paper for the actual conditions of the test being used. The values given by the curves in Fig. 1 of the paper may be used as approximations with any value of divider low side but the divider low side should be between 5 and 10 ohms for a required accuracy of less than 2%.

The Influence of Water Resistivity and Precipitation Rate Upon 60-Cycle Wet Flashover Voltage

AIEE COMMITTEE REPORT

THE WORKING GROUP on Rain Tests was appointed in January 1957, for the purpose of studying problems associated with wet flashover and wet holding tests on high-voltage insulators such as are used on transmission and distribution lines. These problems are intimately concerned with the resistivity of the water used in the tests and the rate of precipitation with which it falls on the test specimen. In connection with the latter, a study of suitable methods for applying the water to the test specimen may well be an area for investigation, particularly for use in wet flashover or holding tests on extra high voltage insulation.

Present European standards for wet testing specify a water resistivity of 10,000 ohms/cm^3 (per centimeter cube) or 3,900 ohms/inch^3 and a precipitation rate of 3 millimeters/minute (0.118 inch/minute), while present American standards specify a water resistivity of 7,000 ohms/inch^3 and a precipitation rate of 0.2 inch/minute.

Wet flashover and wet holding voltages have been specified for practically all classes of high-voltage equipment including transmission and distribution line insulators, station bus insulators, switchgear, circuit breakers, bushings, and lightning arresters, both in America and in Europe. In each area, these values have been determined in terms of the applicable water resistivity and rate of precipitation standards.

In contemplating an international standard for rain tests, which would be applicable both in America and Europe,

---

Paper 58-27, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 15, 1957; made available for printing October 31, 1957.

The personnel of the Working Group on Rain Tests of the AIEE Subcommittee on Techniques for Dielectric Tests are C. J. Miller, Jr., *chairman*, T. L. Dyer, Jr., H. E. Fiegel, B. B. Kingsbury, and D. L. Whitehead. The material for this paper was accumulated mostly in 1938-1940 by the General Electric Company, The Ohio Brass Company, and the Westinghouse Electric Corporation.

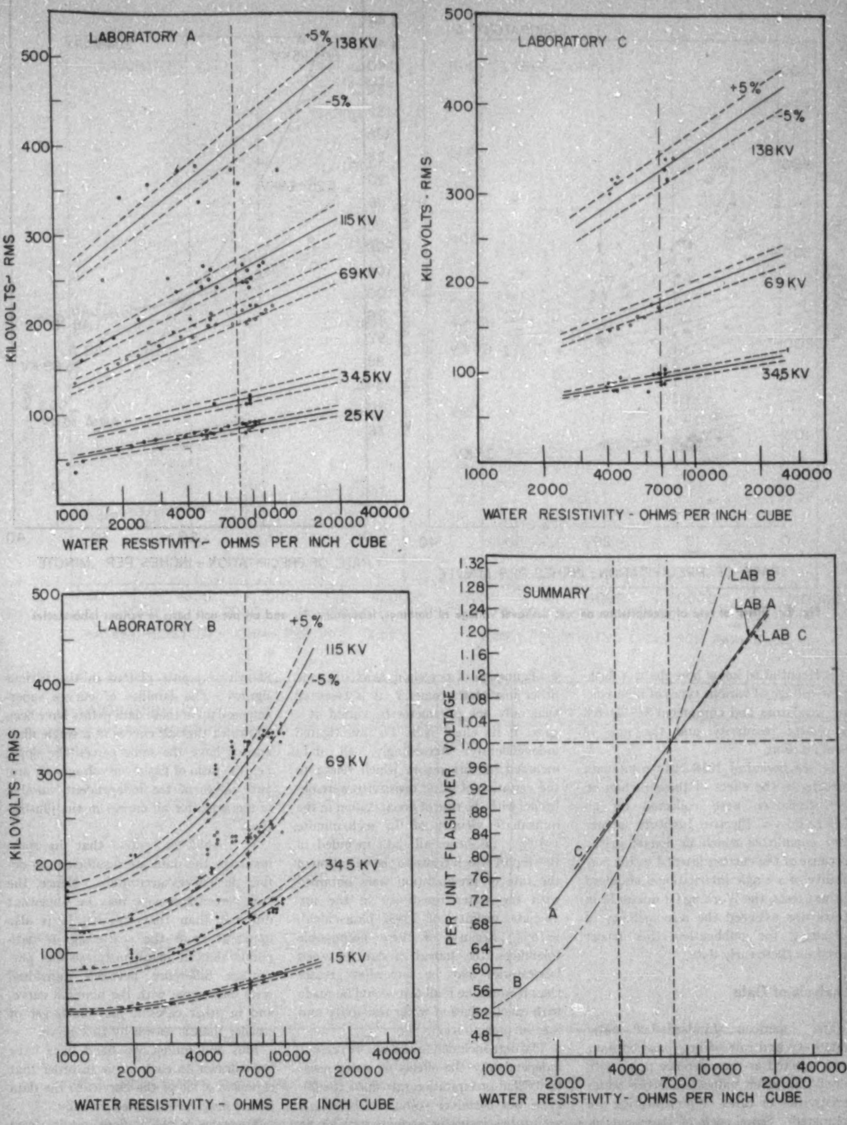


Fig. 1. Effect of water resistivity on wet flashover voltage of bushings in laboratories A, B, and C, and on per-unit basis in various laboratories, shown in summary form

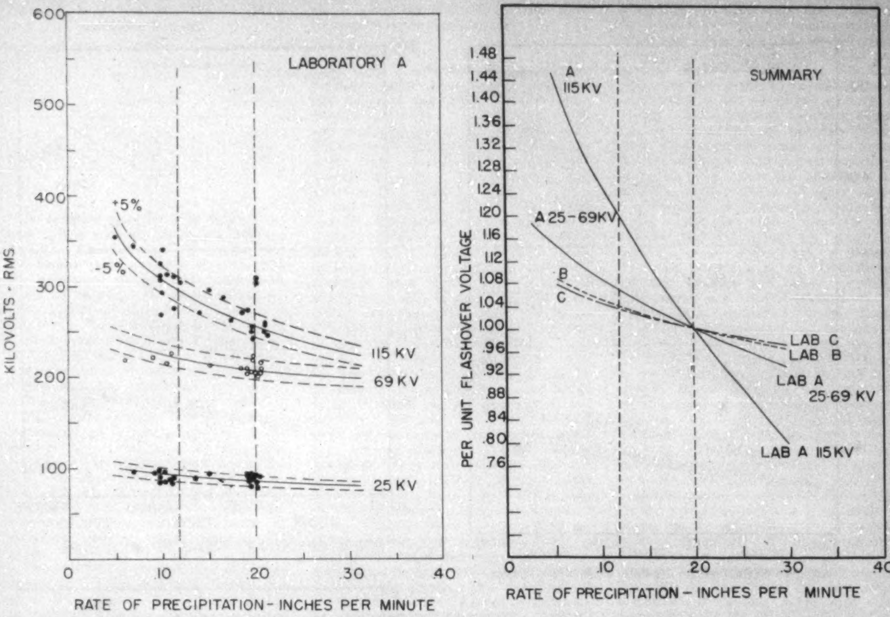


Fig. 2. Effect of rate of precipitation on wet flashover voltage of bushings, laboratory A; and on per-unit basis in various laboratories

it is essential to know how the wet flashover voltage of various types of high-voltate insulators and equipment is affected by water resistivity and the rate of precipitation.

In the period of 1938–1940 some data relating to the effect of these factors on wet flashover were collected by an AIEE–Edison Electric Institute laboratory committee which no longer exists. Because of the current interest in the possibility of a single international standard for wet tests, the Working Group on Rain Tests was assigned the responsibility of preparing for publication this report based on these early data.

Analysis of Data

The American standards of water resistivity and rate of precipitation have been selected as the reference point with which flashover values at other water resistivities or rates of precipitation are compared. Since each of these parameters has an effect upon the wet flashover voltage to a varying degree, depending on such factors as voltage rating, configuration of porcelain, skirt size, and other insulator geometry, it is essential that only one parameter be varied at a time, if its effect is to be investigated independently. Accordingly, all data included in this report which relate to the variation of water resistivity were obtained with the rate of precipitation in the immediate vicinity of 0.2 inch/minute ±10%. Likewise, all data included in this report which relate to the variation of the rate of precipitation were obtained with the water resistivity in the immediate vicinity of 7,000 ohms/inch^3, ±15%. Because of these permissible tolerances, the spread of data between laboratories may be somewhat greater than it would be if all tests could be made with exact values of water resistivity and rate of precipitation.

The data included in this report present independently the effects of water resistivity and precipitation rate upon the 60-cycle wet flashover voltage of bushings, suspension insulator strings, and single units such as pintype, switch, aud suspension insulators.

All of the actual test data included are shown as points plotted in the various figures. The families of curves superimposed upon these data points have been so drawn that all curves in a single illustration have the same percentage slope; i.e., the ratio of flashover voltages for any two values of the independent variable is the same for all curves in the illustration.

It should be noted that in some instances the data are insufficient to define the curves accurately. Hence, the true percentage slope may be somewhat different than that shown. It is also apparent from the scattering of data points that in many instances the percentage difference between individual tests may agree with the nominal curve, and in other cases it may be larger or smaller than is shown by this curve.

Plus and minus 5% band lines have been drawn on each curve in order that closeness of fit of the curves to the data points may be evaluated visually.

To compare the various results from the various laboratories with each other, and to show how the variable parameter influences wet flashover voltages, sum-

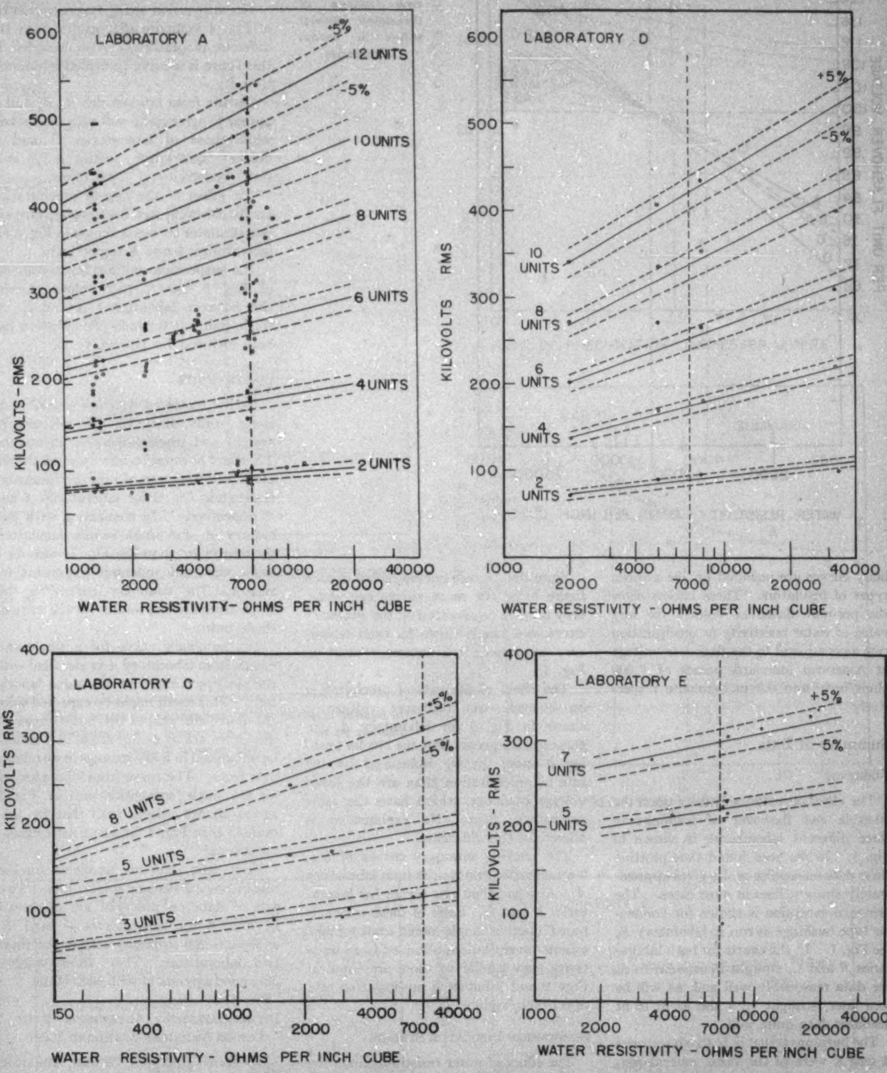


Fig. 3. Effect of water resistivity on wet flashover voltage of suspension insulator strings, laboratories A, C, D, and E

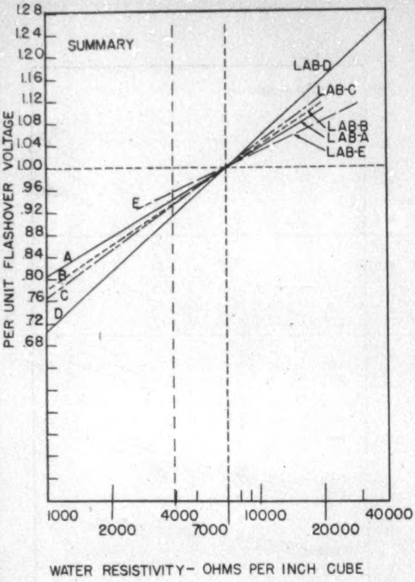


Fig. 4. Effect of water resistivity on per-unit wet flashover voltage of suspension insulator strings in various laboratories

mary curves are included for the various types of insulators. These curves show the per-unit flashover voltage for any value of water resistivity or precipitation rate as compared to the flashover voltage at American standard values of 7,000 ohms/inch^3 and 0.2 inch/minute respectively.

Summary of Data

Bushings

The effect of water resistivity upon the 60-cycle wet flashover of bushings in three different laboratories is shown in Fig. 1. It has been found that plotting these data on semilog scales gives approximately straight lines in most cases. The principal exception is shown for condenser type bushings as run in laboratory *B*; see Fig. 1. In the charts for both laboratories *B* and *C*, straight lines seem to fit the data reasonably well and, as will be seen later, straight lines also seem to fit insulator data quite well.

The bushings tested in both laboratories *A* and *C* were of the same general construction, being of the oil-filled type in which the insulation consisted of a number of concentric barriers of solid dielectric material with layers of oil between adjacent barriers.

Since the average curves shown in each figure have the same percentage slope, they may be represented on the summary curve as a single curve for each laboratory concerned; see summary chart in Fig. 1.

The effect of the rate of precipitation on 60-cycle wet flashover voltage is shown in Fig. 2 for laboratory *A*. In this case, it appears that the 115-kv bushing is more greatly influenced by the rate of precipitation than are the lower voltage bushings, which have the same percentage slope. No explanation is offered for this difference.

The bushing summary curves of Fig. 2 compare the two results from laboratory *A*. Also included are curves for laboratories *B* and *C*. Each of these curves is based upon a single point at 0.1 inch/minute precipitation, obtained from resistivity tests similar to those presented in Figs. 2 and 3 but at a precipitation rate of 0.1 inch/minute.

Suspension Insulator Strings

The effect of water resistivity upon the 60-cycle wet flashover of strings of suspension insulators in the vertical position is shown in Fig. 3 for laboratories *A*, *C*, *D*, and *E* respectively. In each case, the straight lines, which on the semilog plot appear to fit the data quite well, are so drawn that they have the same percentage slope.

The suspension string summary curves of Fig. 4 compare the results from the different laboratories. Also included in this figure is a curve furnished by laboratory *B*.

Results from laboratories *A*, *B*, and *C* appear to agree quite well with each other, while those of laboratories *D* and *E* bracket these curves within ±3% at a water resistivity of 4,000 ohms.

The effect of the rate of precipitation on the 60-cycle wet flashover of suspension insulator strings is shown in Fig. 5 for laboratories *A* and *E* respectively.

The suspension string summary curves of Fig. 5 show a very close correlation between laboratories *A* and *E*. A single point at 0.1-inch precipitation has been included for laboratory *C*.

Single Units

Rather scattered data are available on single units such as pintype, single-switch, and single-suspension insulators. The effect of water resistivity upon the 60-cycle wet flashover of pintype insulators is shown in Fig. 6 for laboratories *A* and *E* respectively. In connection with laboratory *A*, for single-switch insulators, the results are shown in Fig. 7; also for a single vertically mounted suspension insulator. The summary curves in this figure compare the results for the various single units.

The summary curve for a single-unit switch from laboratory *A* is identical with the pintype curve from the same laboratory. This result might be expected since the porcelain shapes on both types of insulators are so nearly alike, and should be influenced in a like manner in wet flashover tests. The curve from laboratory *A* for the single suspension unit in Fig. 7 agrees almost exactly with that for laboratory *D* in Fig. 4 for suspension insulator strings.

Very little data are available on the rate of precipitation on single units. Two sets of data on pintypes are presented in Fig. 8 for laboratories *A* and *E*, as well as the summary curves for these two laboratories. They show reasonably good agreement with each other.

Interlaboratory Agreement with Use of American Standard Rain

To show how well various American high-voltage laboratories agree when present American standards for rain testing are used, the wet flashover voltages for four different string lengths of suspension insulators, as obtained in four

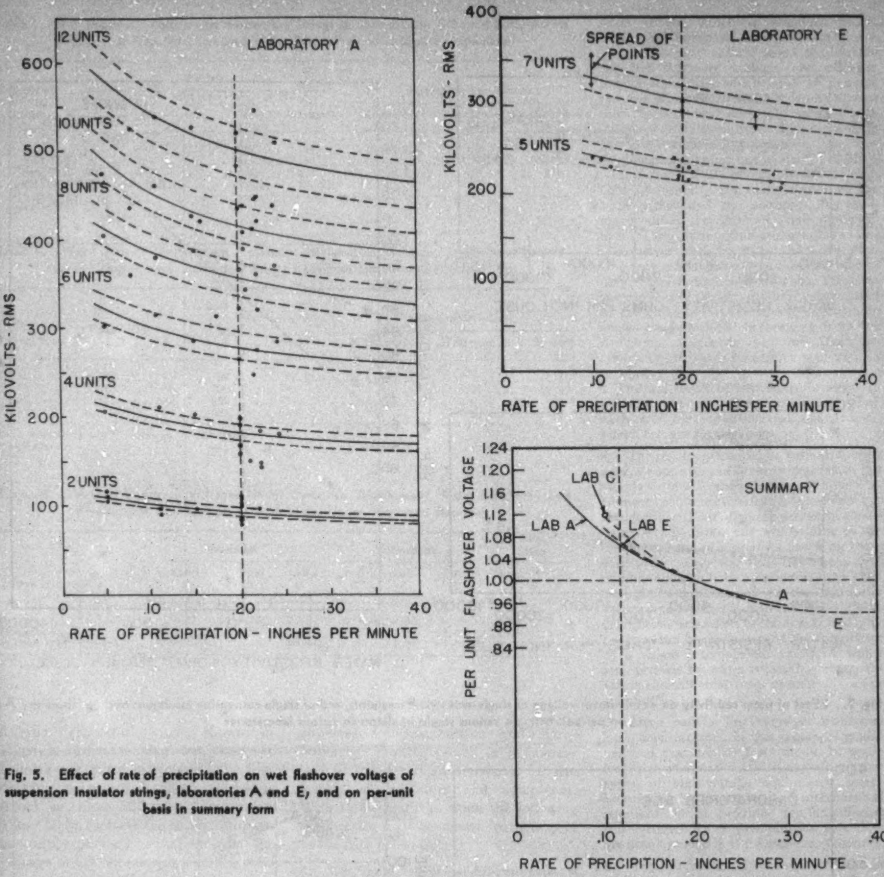


Fig. 5. Effect of rate of precipitation on wet flashover voltage of suspension insulator strings, laboratories A and E, and on per-unit basis in summary form

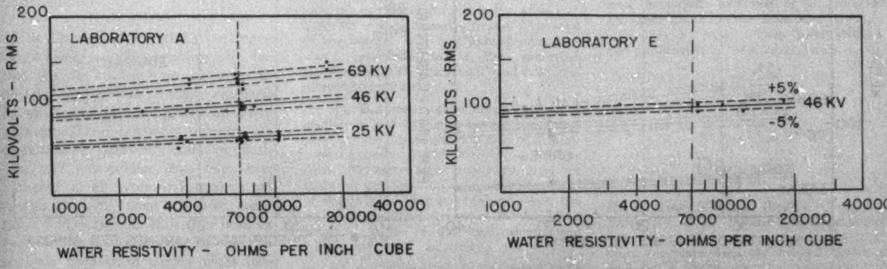


Fig. 6. Effect of water resistivity on wet flashover voltage of pintype insulators, laboratories A and E

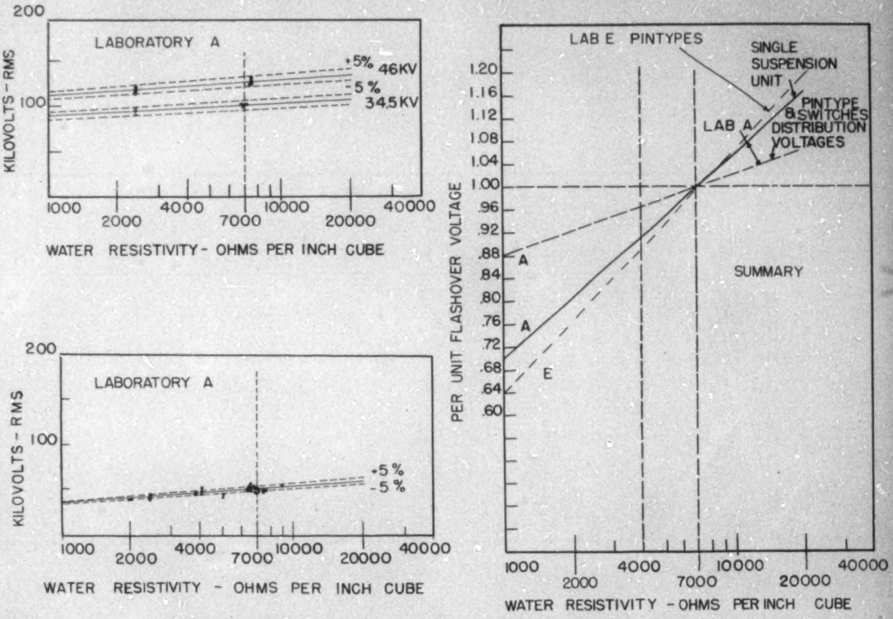


Fig. 7. Effect of water resistivity on wet flashover voltage of single-unit switch insulators, and of single-suspension insulators, both in laboratory A, and on per-unit basis on various single insulators in various laboratories

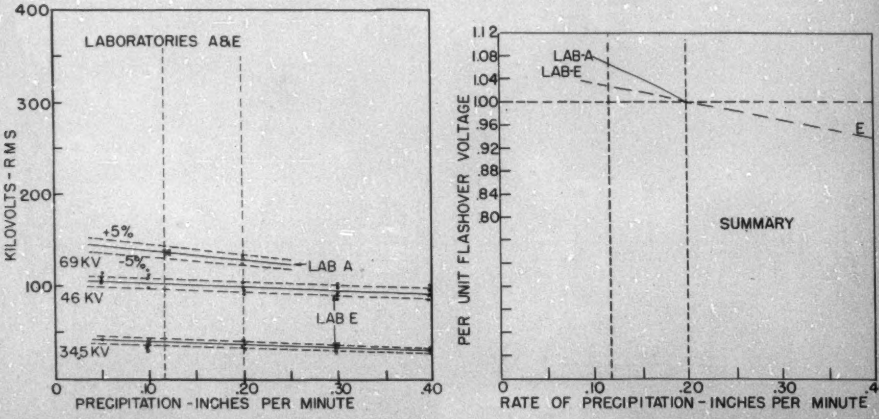


Fig. 8. Effect of rate of precipitation on wet flashover voltage of pintype insulators in laboratories A and E, and on per-unit basis in summary form

Table I. Wet Flashover Voltages on Suspension Strings as Measured in Different Laboratories With Use of American Standard Water Resistivity and Rate of Precipitation

Number of Units in String	Kv, Rms					Deviation, Per Cent		
	Lab* A	Lab B	Lab C	Lab E	Average	Minimum	Maximum	Spread
3	141	133	138	126	134	6.0	5.2	11.2
6	279	254	252	255	260	3.0	7.3	10.3
10	445	413	407	419	426	4.5	4.5	9.0
15	600	599		598	509	0	0	0

* Laboratory.

Table II. Wet Flashover Voltages on Apparatus Insulators as Measured in Different Laboratories Using American Standard Water Resistivity and Rate of Precipitation

Rating, Kv	Kv, Rms				Deviation, Per Cent		
	Lab* A	Lab B	Lab E	Average	Minimum	Maximum	Spread
34.5	101	95	97.5	97.8	3.0	3.3	6.3
115	312	280	290	294	4.8	5.8	10.6
230	582	530	550	554	4.3	5.0	9.3
287	675	590	625	630	6.3	7.1	13.4

Each laboratory tested stacks of its own manufacture.
* Laboratory.

Table III. Per-Unit Flashover Voltages Under European Standards of Water Resistivity and Precipitation Rate Compared With Those Under American Standards

Type of Insulator	Bushings			Suspensions			Pintypes	
	Lab* A	Lab B	Lab C	Lab A	Lab C	Lab E	Lab A	Lab E
Resistivity	0.87	0.77	0.88	0.94	0.93	0.96	0.96	0.89
Precipitation	1.07–1.20	1.05	1.04	1.06	1.09	1.07	1.06	1.03
Product of two factors	0.93–1.04	0.80	0.92	1.00	1.01	1.03	1.02	1.92

* Laboratory.

different laboratories, are shown in Table I. Corresponding wet flashover voltages are given in Table II for four different apparatus insulator stacks as tested in three different laboratories. These tables indicate the maximum plus-and-minus per-cent deviation from the average of all laboratories, and the total per-cent spread in data.

Conclusions

The data included in this report cover only a few types of high-voltage insulators that may be used by the electrical industry. The present American standards for water resistivity and precipitation rate were established a number of years ago, and since that time considerable wet flashover and wet holding data have been gathered on all types of insulators and equipment using these standard conditions, and very little effort has been expended to determine the effects which departures from these values might have upon flashover voltage. Thus, the entire American practice of high-voltage engineering for transmission and distribution of electric power is intimately tied to the American standards for wet testing of insulators and equipment. With this thought in mind the following conclusions may be drawn with respect to the data presented:

1. Wet flashover tests on different types of insulators and bushings using water having the European standard resistivity of 10,000 ohms/cm^3 (3,940 ohms/inch^3), but with 0.2 inch/minute precipitation rate, give wet flashover voltages ranging from 0.77 to 0.96 of the voltage that is obtained when the water resistivity meets the American standard of 7,000 ohms/inch^3 and 0.2 inch/minute precipitation rate.

2. Wet flashover tests on different types of insulators and bushings having the European standard precipitation rate of 3 millimeters/minute (0.118 inch/minute), but with 7,000 ohms/inch^3 resistivity, give wet flashover voltages ranging from 1.03 to 1.20 of the voltage that is obtained when the water precipitation rate meets the American standard value of 0.2 inch/minute and 7,000 ohms/inch^3 resistivity.

3. All results presented in this report show the influence of varying only one quantity at a time, while keeping the other quantity fixed at the American standard value. Since the European standard specifies a lower resistivity, which tends to decrease the wet flashover voltage, and a lower precipitation rate, which tends to increase the wet flashover voltage, it is anticipated that the combined effect of changing both variables simultaneously from American to European standards would have less effect than changing either one alone.

In Table III, the per-unit value for each type of insulator for each laboratory has been taken from the corresponding summary curves for resistivity and precipitation. Only those values are included in this table where both resistivity and precipitation data are available from the same laboratory. The bottom row in Table III gives the product of the two per-unit values listed in the two rows above it, and presumably shows the net effect of changing from the American standards to the European standards of water resistivity and rate of precipitation.

The net result of this analysis seems to indicate that the two factors tend to offset each other for suspension and pintype insulators, but not for bushings.

4. The results presented in Table III, however, do not necessarily mean that European standards of water resistivity and precipitation rate could arbitrarily replace American standards without affecting presently established wet flashover and wet holding values on these types of insulators based on these standards. Very few tests have been conducted in which both parameters have been changed simultaneously to the European values, and hence it is believed that suitable correlation tests are necessary before any progress can be made toward the acceptance of a single international standard which may provide for water resistivity or precipitation rate different from present American standards.

5. Since only a few types of insulators have been included in this report, it is also believed that similar tests should be made on other types of insulators, as well as other types of high-voltage equipment, before attempting to agree on an international standard which provides for a different water resistivity or precipitation rate than are now specified in the American standards.

6. The excellent agreement shown in Tables I and II for 60-cycle wet flashover values obtained in various high-voltage laboratories, when American standards for water resistivity and rate of precipitation are used, indicate the adequacy of these standards, as well as their interpretation and application in these laboratories.

◆

Discussion

H. A. Frey (General Electric Company, Baltimore, Md.): As the committee points out in its timely report, it is apparent that more complete data are needed before a decision is made on the desirability of changing to the International Electrotechnical Commission (IEC) test method; and that the present American standard test does yield remarkably reproducible results when applied in different laboratories.

In considering wet flashover tests and wet withstand tests it must be remembered that their function is to provide standards of measurement which are controllable and reproducible from time to time and from one laboratory to another. They are not intended to be representative of service conditions.

The resistivity and rate of precipitation of natural rain are highly variable. The actual effective resistivity of natural rain on insulator surfaces will probably be more influenced by contamination on those surfaces than by the resistivity of the rain itself. Furthermore, the direction of natural rain may be vertical or at any angle from vertical, depending upon the wind. The size of rain drops may also vary over a wide range. All of these variables will materially influence the wet flashover voltage of insulators in practical service applications.

It is important that our standard tests be made under controlled and reproducible conditions. From such standardized measurements judgment and experience may be applied in establishing insulation which will perform satisfactorily under variable service conditions. If a change in the standards of measurement should result in differences in the standard test results, much of the experience of the past may be lost, unless reliable correction factors are known. It is therefore important that the effect of any change in standards be known before such change is made.

On the other hand, there are some potential gains to be made from a change to the proposed IEC standards for wet tests. The lower precipitation rate reduces the quantity of water which must be stored and disposed of in the laboratory. The lower resistivity may in many cases eliminate the need for preparing water of special resistivity for tests. In many communities the city water supply has a resistivity within tolerable range of the IEC standard of 3,900 ohms/inch^3. These advantages, plus the need for acceptable international standards, make a thorough study of the effect of the IEC method on the wet flashover voltage of various types of apparatus highly desirable.

C. J. Miller, Jr.: As Mr. Frey points out, wet testing standards are intended to provide a method of measurement which is controllable and reproducible from time to time and from one laboratory to another, and are not intended to be representative of service conditions.

It is perhaps unfortunate that American Standards Association (ASA) and IEC standards prescribe different rates of precipitation and water resistivity, and ultimately it is hoped that some means may be found to bring the two standards together.

Up to the present, very few tests have been run in which both rate of precipitation and water resistivity have been modified simultaneously from the ASA values to the IEC values. Accordingly, a tentative test program has been prepared by the AIEE Working Group on Rain Tests in which certain selected specimens representing a wide range of high-voltage apparatus are to be subjected to wet flashover tests under ASA conditions, IEC conditions, and a set of compromise conditions which provide for a water resistivity and rate of precipitation which are intermediate between those values specified by ASA and IEC. When these tests are completed, a single international rain test may be nearer than it appears at present.

An Analog Computer Study of the Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems

OLLE I. ELGERD
MEMBER AIEE

THE most significant change that has taken place in the field of control in recent years has been the appearance of the digital component. The ability of the digital computer to handle and process data quickly and accurately has been the main reason for incorporating digital devices into modern high-performance control systems. The exceptional requirements demanded from today's control systems in the defense area have prompted the development of a large variety of digital components for use in closed-loop control systems for defense applications.

The industry is right now in the process of incorporating many of these components in the industrial control field, and at the same time developing new components for exclusive use in special industrial applications.

This rapid increase in the employment of digital control techniques is forcing the control engineers to take a fresh look at the basic problems associated with this type of control system. The digital components introduce stability problems not encountered in continuous control systems, and special performance criteria apply. The present study is intended to augment the knowledge of digitized control systems. It has been considered of particular interest to investigate the influence of sampling rate on transient system behavior and stability.

The control system of which the digital component is a part can, at least in a first approximation, be considered linear and thus be represented by a transfer function, the complexity of which reflects the complexity of the system. The probability that two systems will have the same transfer functions is rather small, and one realizes therefore immediately the necessity of classifying different control systems into groups characterized by approximately equal performance data. Such a basis for classification has been introduced and made use of throughout the report.

From an analytical standpoint, the

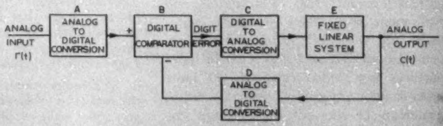


Fig. 1. Schematic "black box" diagram of the general type of system subject to the present investigation

---

Paper 58-31, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted August 29, 1957; made available for printing November 1, 1957.

Olle I. Elgerd is with the University of Florida, Gainesville, Fla.

This report summarizes the results of a study performed at the Advance Engineering Department of General Electric Division of Industry Control, Roanoke, Va. The author wishes to express his appreciation to Dr. F. M. Bailey for valuable suggestions and fruitful discussions.

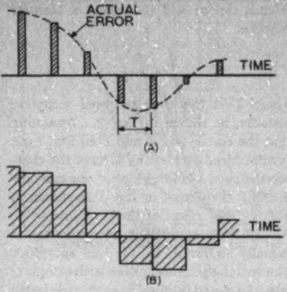


Fig. 2 (left). Graphs illustrating the error signal as it appears at different points in the control system

A—Graph (the envelope) of the actual continuous error as defined by $e(t)=r(t)-c(t)$. The upper train of pulses represents the input $e^*(t)$ to the holding circuit
B—Graph depicts the output from the holding circuit $e_o^*(t)$

digitizer represents a tremendous increase in the degree of computational difficulty. The reason for this is, of course, a consequence of the fact that while a continuous system is governed by a linear differential equation, the digitized system obeys a difference equation. As a result only the very simplest types of sampled-data systems will readily lend themselves to numerical investigations. The analyst has at least a number of time-saving techniques at his disposal. Truxal[1] gave in 1955 in his book a good résumé of the accumulated knowledge of sampled-data systems up to that time, and the field has not been subject to any earth-shaking news since then. Thus, the so-called *z*-transform as described by Ragazzini and Zadeh[3] in 1952 and "Linvill's Method"[8] represent the best analytical tools available today.

The *z*-transform is in a way quite limited in that it describes the control system in a discreet manner, i.e., it gives information about the output only at the sampling instants. This might in many cases be sufficient; in particular is this true when the sampling frequency is relatively high. Linvill's method on the other hand describes all signals as continuous time functions, i.e., it gives information about the system response for any value of time. It is generally known that whatever method is employed, an exorbitant amount of numerical labor is necessary in order to obtain the transient response for a sampled-data system characterized by some degree of complexity. This is very unfortunate and has had the quite understandable consequence that very little published information is presently available.

Objectives of Present Research

The purpose of this investigation is to augment our present knowledge of sampled-data system response. Particular stress has been placed on a study of the transient response to a standard input function (step input). The response has been investigated as a function of the sampling frequency and stability limits have been established in each particular case.

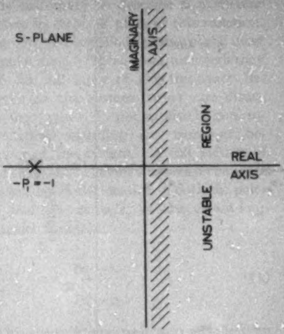


Fig. 5. Case 1: Closed-loop pole configuration

The investigation has been limited to serial-type digital data-handling systems. Considering the large variety of digital control systems in use, a study of present types must of necessity be limited in scope.

One of the main features of this report is its intended applicability of its many diagrams to other control systems. Indeed this was one of the main objectives when the present data were collected and organized. How this has been accomplished and how the diagrams should be interpreted will be covered in more detail later. It will at this time suffice to mention that a classification of different control systems has been made on the basis of location of closed-loop poles and zeros. It will be shown that this is the natural basis for classification and, indeed, the only possible one in a study of present generalized nature.

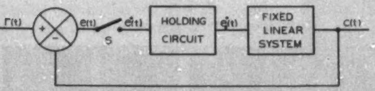


Fig. 3. The components A, B, and D in Fig. 1 can be schematically represented by the "sampler" S as described in the text

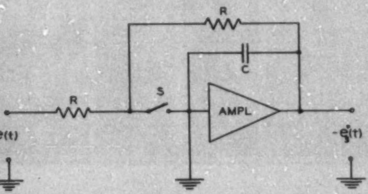


Fig. 4. The sampler plus the holding circuit are simulated by this circuit

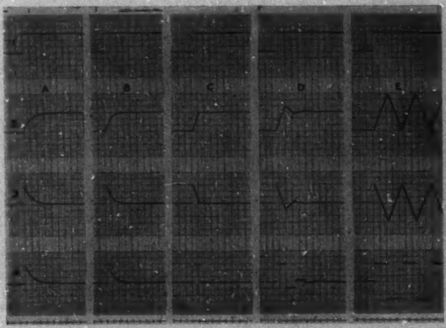


Fig. 6. Case 1: System response to a step input. Chart A depicts an unsampled system, B to E a sampled system. Sampling rate varied

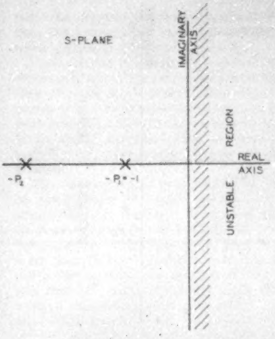


Fig. 7. Case 2: Closed-loop configuration

The afore-mentioned numerical difficulties have been circumvented by adopting analog-computer techniques. All the systems treated in this report were simulated on a Goodyear Electronic Differential Analyzer, and all the data in the report were obtained experimentally.

System Description

It should, from the outset, be clearly pointed out that this study concerns sampled-data systems incorporating a holding circuit. Fig. 1 depicts a prototype of the control system under investigation. The input, instead of being of the continuous type, as shown, can as well be digital, as is the case in programmed card controllers. It is obvious that quite a variety of control systems may be identified with the prototype in Fig. 1. One realizes immediately the necessity of defining at the outset the function of each of the system components. The practical value of the results obtained from a simulation of such a system is a function of how well the assumed operation agrees with the function of the actual hardware.

A large variety of mechanical and electric analog-to-digital converters are presently available.[4–6] They are all characterized by their ability to sample a continuous signal at preset discreet sampling instants. The combined operation of the components A, B, and D in Fig. 1 will thus be to sample the continuous input and output signal $r(t)$ and $c(t)$ at the same instants, compare them and then deliver an error signal consisting of equidistant pulses. In reality, these pulses will have finite duration, but it is general practice to assume that the duration is of infinitesimal duration. This pulsed-error signal is not of suitable wave form to actuate directly the rest of the system and, for this reason, it is fed into the digital-to-analog converter C. (It is worthwhile to mention that in many cases the components A, B, C, and D are physically indivisible.) The converter C is simply a pulse-shaping network, the purpose of which is to extract the usable information from the pulse train. As this information is hidden in the envelope of the pulse train, C must in effect be a low-pass filter. The ideal network should be able to anticipate during the interval between two pulses the magnitude of the next pulse, but as this is extremely difficult or maybe impossible to accomplish, the next best thing is to let the error signal in each interval equal the last preceding pulse. This arrangement is referred to as a holding circuit. Such a device will thus have a transfer function

$$\frac{1-e^{-sT}}{s}$$

where T is the time between sampling instants, as shown in Fig. 2. Assuming that the converters A and D in Fig. 1 are synchronized and letting C have the characteristics of a holding circuit one can now simplify the circuit in Fig. 1 as shown in Fig. 3. The effect of the components A, B, and D is equivalent of the switch S (usually referred to as the sampler). The switch opens and closes with sampling rate, the closed interval in each cycle being infinitesimally short.

Simulation Techniques

The simulation of the fixed linear part of the system in Fig. 3 does not impose any particular problems, but follows standard procedures.[7] It might be of interest, however, to describe in some detail the method employed in simulating the sampler plus the holding circuit. The circuit depicted in Fig. 4 was found to do a very satisfactory job.

It has the advantage of permitting us to use one of the computer amplifiers. If the error signal $e(t)$ is applied to the input, this circuit will deliver with very good accuracy the desired sampled signal $e_s^*(t)$ (with reversed sign). For proper functioning, the following conditions must be fulfilled.

1. The switch S must have a closing cycle similar to the sampler, i.e., it should be closed during a very short interval of the cycle.
2. The time-constant RC should be much

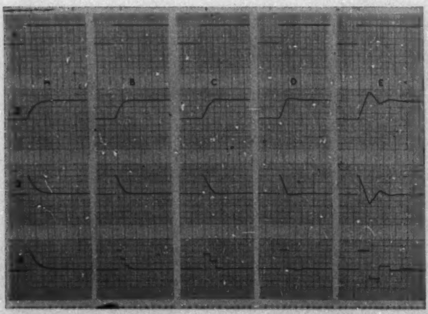


Fig. 8. Case 2: System response to a step input. Chart A depicts an unsampled system, B to E sampled systems. Sampling rate varied. $P_2=8$

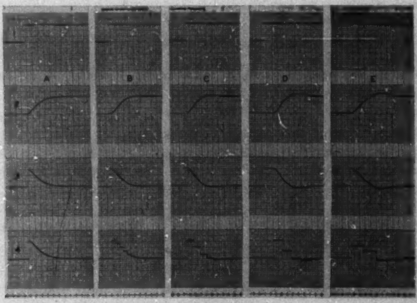


Fig. 9. Case 2: System response to a step input. Chart A depicts an unsampled system, B to E sampled systems. Sampling rate varied. $P_2=1$

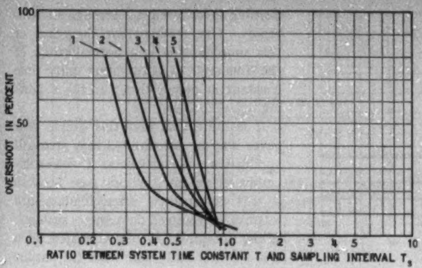


Fig. 10. Case 2: Overshoot versus sampling rate

The different plots apply to systems having:

1. $p_2=1$
2. $p_2=2$
3. $p_2=4$
4. $p_2=8$
5. $p_2=\infty$

less than any other time constants in the circuit in which it is operated.

If these conditions are fulfilled, the output voltage will, when S is closed, very quickly assume the same value as the input (with reversed sign). When S then opens, the output voltage due to the action of the capacitor will remain constant until S closes next time. The switch S was designed as a rotating commutator driven by a d-c motor thus permitting a convenient variation of the sampling frequency. It should be noted that the closing time of S must not be too short; the capacitor must be given ample time to assume its correct value of potential. Evidently this condition will restrict the sampling rate. The following parameter values were found suitable:

$R=0.01$ megohm
$C=0.5$ microfarad
Closing time of switch equals $^1/_{25}$ of period time T.

The preceding values give a time-constant of 5 milliseconds. If we require that the switch should be closed at least during 25 milliseconds, (this will allow the capacitor to charge up to 99.5% of the input voltage) the maximum permissible sampling rate will be:

$$\frac{1}{25(0.025)}=1.60 \text{ cycles per second}$$

A relatively slow computer time scale must therefore be chosen.

Classification of Control Systems on Basis of the Closed-Loop Pole and Zero Locations

This study is in effect a comparison of the response between a sampled and an unsampled or continuous system. It should be noted in passing that a continuous system can be considered to be a special case of a sampled system; one having infinite sampling frequency. As a continuous system is uniquely determined from its open-loop transfer function $G(s)$, it is customary to describe a particular system in terms of plots of this function. The Nyquist and Bode plots are the most popular ones. These plots, although they tell the full story of the frequency behavior of the system, suffer one disadvantage; they do not explicitly give information regarding the transient behavior of the system. A full picture of the transient behavior requires a complete knowledge of the location of both the poles p_ν and the zeros z_ν of the closed-loop transfer function:

$$\frac{G(s)}{1+G(s)}=K_c\frac{\overset{m}{\Pi}(s-z_\nu)}{\overset{n}{\Pi}(s-p_\nu)} \qquad (1)$$

(This study limits itself to true follow-up systems, i.e., systems having zero-position error for a step input. This means that

$$K_c=\frac{\overset{n}{\Pi}(-p_\nu)}{\Pi(-z_\nu)}$$

which may readily be verified by using the "final-value theorem.")

Once the z_ν; s and p_ν; s are known, it is possible to find immediately the transient response:

$$c(t)=\sum_{\nu=1}^{n}A_\nu\cdot e^{p_\nu t} \qquad (2)$$

The amplitudes A_ν are most easily expressed in terms of distances between the poles and zeros[8] or can be analytically

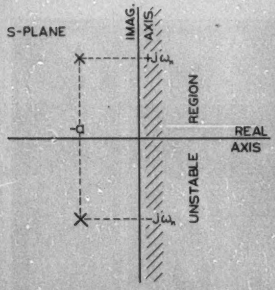


Fig. 11. Case 3: Closed-loop pole configuration

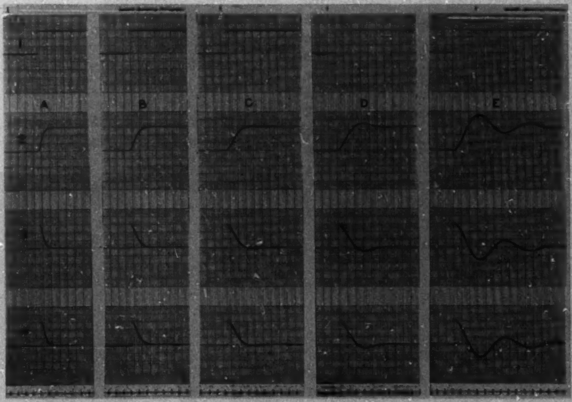


Fig. 12. Case 3: System response to a step Input. System unsampled

A. $a=4$
B. $a=2$
C. $a=1$
D. $a=0.6$
E. $a=0.2$

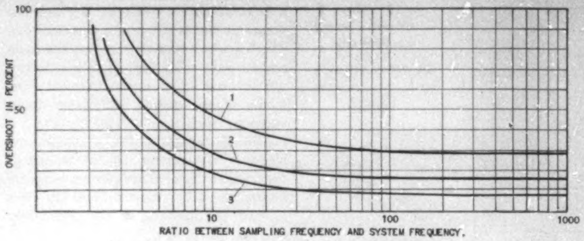


Fig. 13. Case 3: Overshoot versus sampling frequency

The three different plots refer to systems having:

1. a=0.4 2. a=0.6 3. a=0.8

determined by using Heavisides Expansion Theorem.

(The problem of finding the poles and zeros for a given system has been extensively treated by many authors.[9] It should be noted from equation 1 that the closed-loop zeros z_v coincide with the zeros at the open-loop transfer function $G(s)$ but the closed-loop poles p_v must be found by solving the characteristic equation

$$1+G(s)=0$$

This last task is comparatively simple if the powerful root-locus techniques are utilized.)

At a first glance it would seem to be a prodigious task to try to investigate every possible pole-zero combination, of which there exists an infinite number. Fortunately, most control systems, even the most complex ones, can be characterized by a very few poles and zeros, namely the ones located close to the origin in the complex s-plane.[9] We refer to those poles and zeros as the predominant ones. The reason for this predominance follows from equation 2, in which some of the A_v's will be very large compared to others. The actual transient response may thus be very accurately determined by taking only the large-amplitude terms into account.

Basis for Comparison of System Response

One of the classical control system problems is to find a criterion that will give information about the "best" system response. Several criteria have been proposed, for instance the relatively popular "error-square" criterion, but they all require explicit knowledge of the error as a function of time. Considering the fact that among control engineers there exists much doubt as to the relative value of these error criteria, it does not seem to be of any particular value to choose too sophisticated a basis for comparison. The choice has instead been governed by the desire to use a measuring rod which will allow quick reference. The overshoot of the system output serves this purpose well. Not only is the overshoot a measure of the relative stability but it gives the control engineer a good picture of the transient behavior of the system as well.

We will now proceed to investigate different systems characterized by increasingly complex closed-loop pole-zero configurations.

Case 1. Negative Real Pole, No Zero (Nonoscillatory Case). Fig. 5 depicts the simplest possible pole-zero configuration. In this case, the closed-loop transfer function can be written:

$$\frac{C(s)}{R(s)}=\frac{G(s)}{1+G(s)}=\frac{p_1}{s+p_1}=\frac{1}{s+1}$$

and thus:

$$G=\frac{1}{s}$$

(Note that no generality is lost by letting p_1 assume the value unity. This simplification is adhered to throughout this report. One realizes that the effect of this simplification is a normalization of the time scale so as to obtain unity time constant as demonstrated in the following.)

It is quite true that this simple case never may be encountered in a practical application. Nevertheless, it serves superbly as an introductory case because it will demonstrate some fundamental features of sampled-data servo systems. If the unsampled system is subject to a step input of unity amplitude the output will be:

$$c(t)=L^{-1}\left[\frac{1}{s}\cdot\frac{1}{1+s}\right]=1-e^{-t}$$

i.e., a purely nonoscillatory exponential response, as demonstrated in Fig. 6(A).

As to the charts in this report the following general remarks are made. From top to bottom the recordings show:

Channel 1.	System input $r(t)$
Channel 2.	System output $c(t)$
Channel 3.	Unsampled error signal $e(t)$
Channel 4.	Sampled error signal $e_s^*(t)$

In addition each graph contains at the lower edge a one-second time marker.

In Figs. 6(B) to 6(E), the system response is recorded for four different sampling frequencies. The following points of interest should be noted from those charts:

1. Although the system is inherently nonoscillatory, it will for sufficiently low sampling rate become oscillatory as evidenced by Fig. 6(D).
2. If the sampling rate is still decreased, the system will become unstable. Fig. 6(E) depicts the limit case.
3. For a certain sampling rate, the sampled-data system will have a more attractive response than the continuous system; the response of Fig. 6(C) is certainly to be preferred over Fig. 6(A).

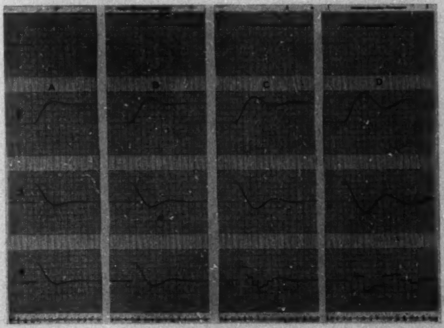


Fig. 14. Case 3: a = 0.6. The charts show the response to a step input if the sampling rate is varied

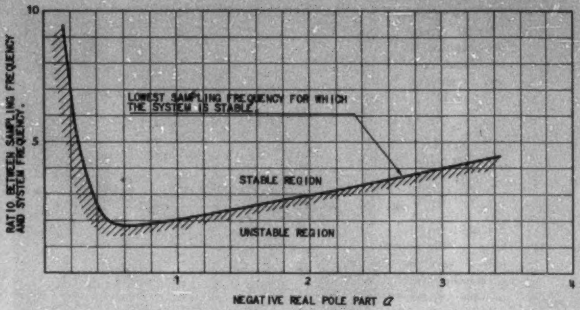


Fig. 15. Case 3: Minimum stable sampling rate versus damping factor a

The first two points apply to all control systems, the last one only to very sluggish ones.

In addition, the one-pole system is the only one where the output will consist of straight line segments.

Case 2. Two Negative Real Poles, No Zero (Nonoscillatory Case). This pole configuration is shown in Fig. 7. The system is characterized by:

$$\frac{G(s)}{1+G(s)}=\frac{p_2}{(s+1)(s+p_2)}$$

and thus:

$$G(s)=\frac{p_2}{s(s+1+p_2)}$$

The system response is built up by two monotonously decaying exponentials and the system is thus nonoscillatory. For increasing magnitude of p_2 this case approaches case 1 as a limit case. For large values of p_2 one can therefore expect a system response similar to the one in case 1. This is illustrated by the graphs of Fig. 8 which refer to a system having $p_2=8$.

For p_2 values slightly larger than unity, the system will in effect behave as if it had a double pole at $s=-1$. The double-pole behavior is demonstrated in Fig. 9, which clearly demonstrates the increased sluggishness. Charts *B*, *C*, *D*, and *E* of Figs. 8 and 9 show the response to a step input for sampled systems.

Fig. 10 finally gives in very condensed form information about the overshoot of the system as a function of the sampling rate. Because the system is inherently nonoscillatory, the sampling rate is expressed as a ratio T/T_s, where T is the largest time constant of the system ($T=1/p_1=1$ second) and T_s is the inverse of the sampling frequency f_s ($T_s=1/f_s$). The different plots refer to p_2 values in the complete range zero to infinity ($0-\infty$). Note that $p_2=\infty$ actually refers to case 1. The plots are interesting in that they reveal that it is possible to reduce the sampling frequency to a value of approximately $1/T$ cps before the system ceases to be nonoscillatory. For a further decrease in sampling frequency, the system turns oscillatory quite abruptly.

Case 3. Two Conjugate Complex Poles, No Zero (Damped Oscillatory Case). This case is very important because a large number of control systems are characterized by a (more or less) dominant complex pair of poles. The system is inherently oscillatory, the natural frequency of the system being ω_n radians per second (see Fig. 11), and the damping being expressed by the damping factor a.

In order to normalize, ω_n is set equal to unity and one then gets:

$$\frac{G(s)}{1+G(s)}=\frac{1+a^2}{(s+a+j)(s+a-j)}=\frac{1+a^2}{(s+a)^2+1}$$

and also:

$$G(s)=\frac{1+a^2}{s(s+2a)}$$

The sequence of charts in Fig. 12 demonstrates clearly what effect a variation of the real part of poles has on the system response. As the poles move to the right in the s-plane and thus approach the imaginary axis, the system gets more and more oscillatory. If the poles move far to the left, the system, although still oscillatory (not discernible on the chart), will approach the double-pole behavior shown earlier in Fig. 9(A).

If a is permitted to vary between 0.4–0.8, the overshoot of the unsampled system will vary between 29–8% and this might therefore be considered to represent the interesting region. The transient response of the sampled system was investigated for three different a values within the above range. The results of the study are condensed into the three curves in Fig. 13, showing the overshoot as a function of sampling frequency. The series of charts in Fig. 14 depict a typical run. (Note that the ratio of sampling frequency to system frequency is chosen as abscissa.)

The most significant information to be derived from these curves is that it takes

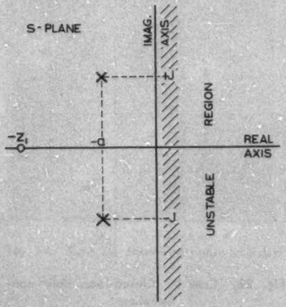


Fig. 16 (left). Case 4: Closed-loop pole-zero configuration

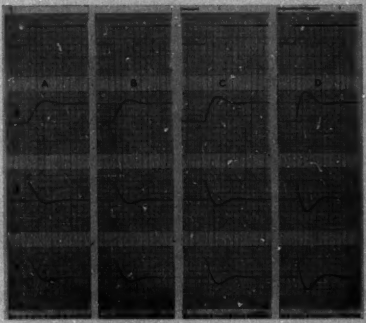


Fig. 17 (right). Case 4: System response to a step input. System unsampled. a=0.6

A. $z_1=4$
B. $z_1=2$
C. $z_1=1$
D. $z_1=0.6$

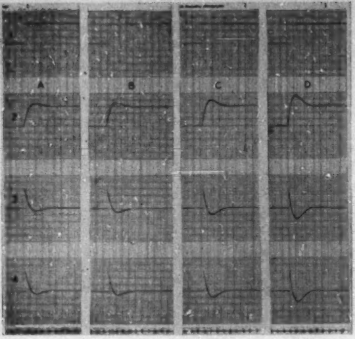


Fig. 18 (left). Case 4: System response to a step input. System sampled. a = 1

A. $z_1=1.2$
B. $z_1=1.0$
C. $z_1=0.8$
D. $z_1=0.6$

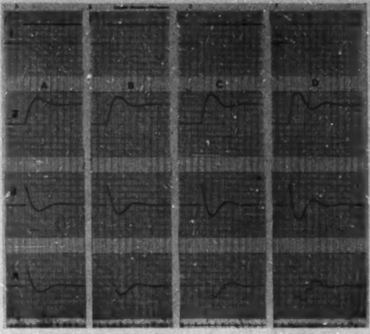


Fig. 19. (right) Case 4: The effect of variation of sampling rate on the system response. $a=0.6$, $z_1=1$

quite a high sampling frequency to preserve the desirable transient characteristics of a system. As an example: A system that unsampled has an overshoot of 15% will have a 40% higher overshoot for a sampling frequency of as high as 25 times the inherent system frequency.

To make the study complete, it was also investigated what effect the sampling has on the system stability. The procedure was to lower the sampling frequency until the system went unstable. The sampling frequency corresponding to this stability limit was plotted as a function of the negative real part a with the result shown in Fig. 15. One should observe the rapid rise below $a=0.5$.

Case 4. Two Conjugate Complex Poles, One Zero. This system, having the pole-zero configuration shown in Fig. 16, is characterized by:

$$\frac{G(s)}{1+G(s)}=\frac{1+a^2}{z_1}\cdot\frac{s+z_1}{(s+a)^2+1}$$

and

$$G(s)=\frac{1+a^2}{z_1}\cdot\frac{s+z_1}{s\left[s+\frac{2az_1-1-a^2}{z_1}\right]}$$

The effect of introducing the zero is most effectively demonstrated by the sequence of charts in Fig. 17, showing four recordings all made with a system having a fixed complex pole-pair ($a=0.6$) but where the zero occupies four different positions ($z_1=4$, 2, 1, and 0.6). When the zero is located far to the left the system approaches the one treated in case 3. This is confirmed by chart *A* which is almost identical to chart 12(D). If the zero moves to the right the rise time decreases, and we obviously obtain a faster system. This improvement will, however, be gained at the cost of a more oscillatory system.

It is interesting to note that by placing the poles relatively far to the left and the zero relatively far to the right one can obtain a rather attractive response. This is demonstrated by Fig. 18, which refers to a system having $a=1$. The four

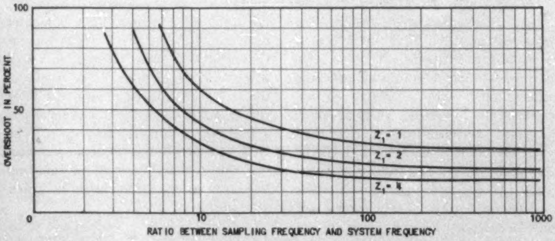


Fig. 20. Case 4: Overshoot versus sampling rate. $a=0.6$. The three plots refer to z_1-values as indicated in the figure

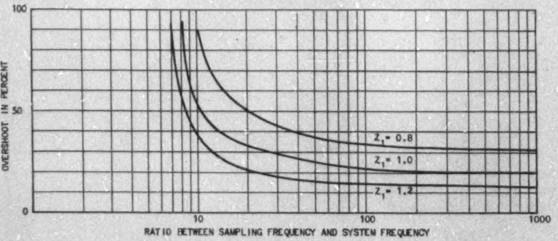


Fig. 21. Case 4: Overshoot versus sampling rate. $a=1.0$. The three plots refer to z_1-values as indicated

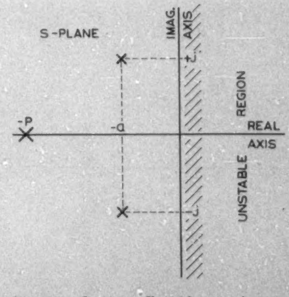


Fig. 22. Case 5: Closed-loop pole configuration

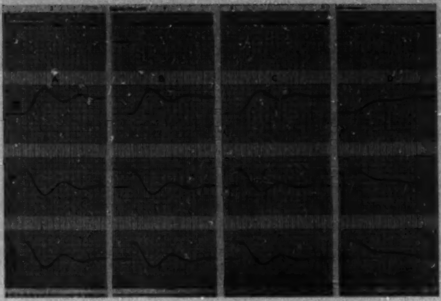


Fig. 23. Case 5: System response to a step input. $a=0.2$

A. $p=4$
B. $p=2$
C. $p=1$
D. $p=0.2$

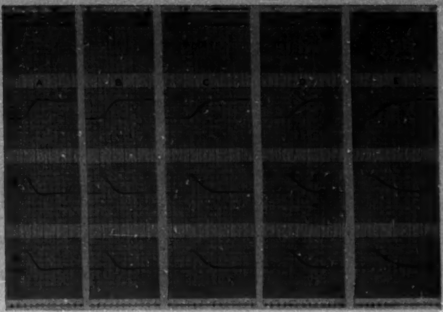


Fig. 25. Case 5: System response to a step input. $a=1.0$

A. $p=4$
B. $p=2$
C. $p=1$
D. $p=0.8$
E. $p=0.5$

charts are depicting the response for $z_1=1.2$, 1.0, 0.8, and 0.6. It should be noted that the location of the zero in this case is very critical; by putting it little too far to the right the system gets violently oscillatory.

If this system is made subject to sampling one must obviously expect more severe stability problems than in case 3. The four charts in Fig. 19 demonstrate this for a system characterized by $a=0.6$ and $z_1=1$. One can easily notice the tendency for large overshoot (and impaired stability) even for relatively high sampling frequencies. Even more revealing in this respect are the diagrams in Figs. 20 and 21. In Fig. 20 the overshoot is plotted versus sampling frequency for a system characterized bv $a=0.6$ and $z_1=1,2$ and 4. Fig. 21 refers to a system having $a=1$ and $z_1=0.8$, 1.0 and 1.2.

Case 5. Two Conjugate Complex and One Negative Real Pole. An additional pole positioned along the negative real axis as shown in Fig. 22 adds one more exponential term to the time response. The system is characterized by:

$$\frac{G(s)}{1+G(s)}=\frac{p(1+a^2)}{(s+p)\cdot[(s+a)^2+1]}$$

and

$$G(s)=\frac{p(1+a^2)}{s[s^2+(p+2a)s+1+a^2+2ap]}$$

This additional pole has the opposite effect as had the added zero in case 4; it will make the system slower and also more stable. The charts in Figs. 23, 24, 25 demonstrate this for three different locations of the complex pole-pair ($a=0.2$, 0.6, and 1.0). One can notice the decrease in overshoot and increase in rise time as the real pole moves to the right. Should the pole move too far to the right in the s-plane the system response will be intolerably slow; see for instance chart 23(D).

This stabilizing effect will also be felt in a sampled case-5 system. Fig. 26 depicts a typical computer run, from which one easily realizes the inherent stability of the system even for very low sampling frequencies. Fig. 27 finally shows how the overshoot varies as a function of sampling frequency for four different representative pole configurations.

Conclusions

The present report contains results from an investigation of the transient characteristics of sampled-data system. The

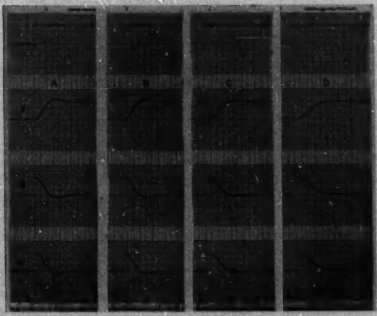


Fig. 24. Case 5: System response to a step input. $a=0.6$

A. $p=4$
B. $p=2$
C. $p=1$
D. $p=0.6$

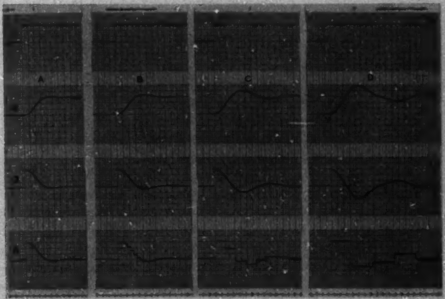


Fig. 26. Case 5: The influence of sampling rate on system response. $a=0.6$, $p=1$

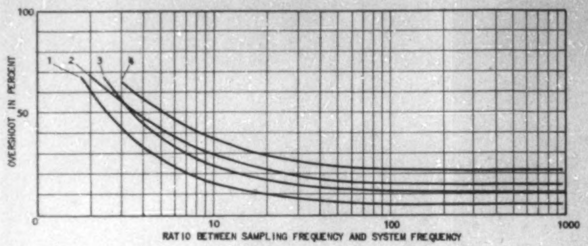


Fig. 27. Case 5: Overshoot versus sampling frequency

The four plots refer to:

1. $a=0.6$ $p=1$
2. $a=0.4$ $p=1$
3. $a=0.6$ $p=2$
4. $a=0.4$ $p=2$

presented data have been found experimentally from simulation of the studied systems on a differential analyzer. In order to optimize the generality of this study and thus make the results applicable to other systems than the specific ones studied herein the main problem from the outset was to classify the systems in an orderly and logical fashion. The author makes no claim of a complete success of this aim but the idea of using the closed-loop zero and pole locations as a basis for classification is indeed both simple and promising and to the author's opinion outclasses any other thinkable basis for a classification.

The method is based on the fact that the large majority of control systems are fundamentally characterized by only a few of its many poles and zeros. Those predominant closed-loop poles and zeros are the ones located closest to the origin in the complex s-plane. Five different basic-pole zero configurations have been included in this study. Transient response and particularly the influence of sampling rate on system stability have been studied.

It is anticipated that the data made available in this report will be easily digested by the increasing number of control engineers who utilize the powerful root-locus design techniques.

Although the results are not directly applicable to systems for which only the frequency response is known (for instance in the form of a Bode plot) they still can be made useful as follows:

1. Correlate the Bode plot of the system under study with the transfer functions as given in each of the foregoing five cases. By a straight-line approximation within the frequency range of importance find out what the break frequencies are and also the slope between those frequencies (expressed for instance in decibles per octave). Choose then the case that agrees with the given system. Note, the preceding statement that some of the closed-loop poles and zeros of a system are predominant corresponds to the equally true statement that only the lower region of the frequency response need to be considered when the system response is discussed.

2 Vary then the system parameters (a, z_1, etc.) so that the given Bode plot approximately coincides with the Bode plot of the case that has been chosen.

References

1. Control System Synthesis (book), **John G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, chap. 9.

2. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 225–32.

3. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **William K. Linvill.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1779–88.

4. Digital Methods in Control Systems, **D. F. Nettel.** *Electronic Engineering*, London, England, Mar. 1956, pp. 108–14.

5. A Survey of Analog-to-Digital Converters, **H. E. Burke.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, 1953, p. 1455.

6. Analog-to-Digital Converters, **George G. Bower.** *Control Engineering*, New York, N. Y., Apr. 1957, pp. 107–16.

7. Electronic Analog Computers (book), **G. A. Korn, T. M. Korn.** McGraw-Hill Book Company, Inc., 1952, chap. 3.

8. The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems, **J. H. Mulligan, Jr.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, May 1949, pp. 516–29.

9. Control-System Dynamics (book), **W. R. Evans.** McGraw-Hill Book Company, Inc., 1954.

◆

Discussion

G. Farmanfarma and **F. Mullin** (University of California, Berkeley, Calif.): This paper is a good addition to the present literature in the field of sampled-data systems. Its objectives are carefully stated and carried out. These experimental results provide a clear picture of the transient behavior of sampled-data systems as their sampling rates are varied.

Although the word "earthshaking" may be individually interpreted, we do feel there has been significant contributions in this field since 1955 which should be added to the list of references. The addition of the modified z-transform[1,2] provides a technique for computing the response of sampled-data systems as a continuous function of time. Thus, all of the results presented in this paper are obtainable analytically.

The effect of the sampling period on the transient behavior of sampled-data systems has been mentioned by some authors.[3–7] In reference 6 a sampling rate has been specifically chosen to give a dead-beat response. Furthermore, in this reference the modified z-transform has been used to assure a continuous dead beat response; i.e., a response which is equal to the input for all values of time greater than the specified rise time. Also, we would like to point out that a similar computer study has been carried out in reference 4. However, we feel that many people will find the present paper helpful in the study of sampled-data systems.

REFERENCES

1. Synthesis and Critical Study of Sampled-Data Control Systems, **Eliahu I. Jury.** *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 141–49.

2. The Pulse Transfer Function and Its Application to Sampling Servo Systems, **R. H. Barker.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, 1952, pp. 302–17.

3. Analysis and Synthesis of Sampled-Data Systems, **Eliahu I. Jury.** *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

4. Transient Analysis of Sampled-Data Control Systems. **G. W. Johnson, D. P. Lindorff.** *Ibid.*, pt. II, July 1954, pp. 147–53.

5. Network Compensation of Error-Sampled Feedback Control Systems, **J. Sklansky.** *Technical Report T-7/B*, Columbia University, New York, N. Y., Apr. 1957.

6. Discrete Compensation of Sampled-Data and Continuous Control Systems, **Eliahu I. Jury, William Schroeder.** *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 317–25.

7. Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 427–35.

Philosophy of Applying Digital Computers to the Design of Electric Apparatus

P. A. ABETTI
MEMBER AIEE

W. J. CUTHBERTSON
NONMEMBER AIEE

S. B. WILLIAMS
ASSOCIATE MEMBER AIEE

ONE DEFINITION of the word "philosophy" given by Webster's dictionary is "the body of principles underlying a given branch of learning or major discipline." In this paper the authors attempt to establish the philosophy, according to this definition, of applying digital computers to the design of electric apparatus. Although the authors' experience has been mainly in the field of power transformers, special-purpose transformers, and current-limiting reactors, they have followed closely similar applications to other types of electric apparatus, notably distribution transformers, medium transformers, and rotating machines. Because of this, and in view of the unity of the science and art of electrical design, it is felt that the concepts and conclusions developed in the paper will be generally valid for the broad class of electric apparatus.

An electric machine may be visualized as the site of several superimposed fields, namely, an electric field, a magnetic field, a current field, a field of mechanical forces, a field of conductive and radiated heat flow, a field of elastic forces in the material stressed by magnetic forces, and a field of fluid flow of the cooling medium. These fields must be considered from the steady-state and also from the transient point of view, since practically all machines are subjected more or less frequently to abnormal overstresses, in addition to the normal service stresses. The designer's problem is to analyze and modify these closely interrelated fields so as to translate the specifications into constructional features and data, with the ultimate aims of safety, simplicity, efficiency, and economy. The design of electric apparatus is both a science and an art: a science because it follows established and universally accepted physical and mathematical principles which have been verified by the experimental method; and an art in that knowledge of these principles is often insufficient to produce a correct and economic design, and must be supplemented by considerable judgment and experience which can be acquired only after intensive training on the job. This is due to: 1. the complexity of interrelations appearing in the design process, and 2. the lack of specific knowledge in certain areas which are continuously affected by new technological developments. The application of digital computers to the solution of mathematical, engineering, and data-processing problems may also be considered as both a science and an art for the same reasons.

Both electrical design and computer applications are quite new among the applied sciences. Electrical design is hardly 75 years old, since it followed the discovery of the rotating electric machine and of the transformer in the latter part of the last century. Digital computers, as is well known, were developed for military purposes toward the end of World War II, and only later were applied to nonmilitary computations. Starting less than 10 years ago, digital computers were first applied to the solution of individual mathematical problems arising from the design of electric machines; for instance, the solution of differential equations, the inversion of matrices, the calculation of mechanical stresses in a component, etc. These applications were limited to specific portions or phases of the design problem, which were particularly laborious, urgent, or difficult. Although extremely valuable in practice, these limited applications cannot be defined as electrical design in the broad sense given in the foregoing.

The marriage of the still young electrical design art and the infant digital computer art took place about 5 years ago. At that time, two of the authors applied the IBM (International Business Machines Corporation) card-programmed calculator (CPC) to the entire design process and, by an automatically convergent iterative procedure, were able to determine the constructional features which would meet the specifications.[1,2] The first power transformer designed in this manner came out of the Schenectady CPC in May 1953. Similarly, the application of computers to the design of rotating machines was first discussed in 1954.[3] Since that time, several papers have been written in this country on the subject of applying computers to electrical design, and work is actively progressing in this new and promising field.[4–9] The widespread use of digital computers for scientific, engineering, and data-processing applications in Europe has lagged behind that in the United States. In 1956, a paper[10] on the use of computers for power transformer design was published in Great Britain, which has been ahead of other European countries in the computer field. In a recent trip to Europe, one of the authors found that Swedish, French, German, Belgian, and Italian manufacturers of electric apparatus are investigating or using the digital computer as a design tool, and were well aware of its potentialities and advantages.

Thus, while the science and art of designing electric apparatus with digital computers is still in its infancy, it has, to date, achieved noteworthy practical results and is forging ahead rapidly, in this country and abroad. Therefore it appears opportune at this time to summarize the available knowledge and experience, and to establish a philosophy which may be used as a guide for future developments and applications.

It might appear more logical to establish first a philosophy, and then to apply it. But it is an outstanding characteristic of the American people, and particularly of engineers, that they want to go into action as soon as possible with practical work without worrying too much in this first stage about theoretical agreement on broad principles, philosophical niceties or details, etc. This is the characteristic of a growing, optimistic, aggressive society, goal-oriented rather than principle-oriented. However, sooner or later the time comes to organize resources and systematize knowledge in order to avoid excessive empiricism, lack of direction, and waste of manpower, time, money, and effort.

In this paper the purpose, the principles, and the process of design in electric apparatus will first be reviewed. Then, the characteristics of the modern digital computer will be reviewed. After this, the principles of applying the digital computer to electrical design according to

---

Paper 58-233, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 16, 1957; made available for printing November 20, 1957.

P. A. Abetti, W. J. Cuthbertson, and S. B. Williams are with the General Electric Company, Pittsfield, Mass.

the four steps of design analysis, design synthesis, optimization, and standardization will be developed at length. Wherever possible, examples, taken from the authors' experience, will be given of the practical application of these principles.

Design

Purpose

In its broadest sense, design entails the translation of specifications into information which describes the apparatus in the necessary detail and its performance. Its purpose, then, is to create an apparatus which will satisfy all of the desired specifications; conform to all applicable industry and company standards; be pleasing in appearance; and use all the materials within it and the applied labor, in the most economical fashion. In industry today there are many reasons why a particular design may be required and thousands of designers are employed to meet the needs. A design may be necessary to accomplish the following:

1. Set up a product line, which will satisfy the specifications and characteristics required by the majority of prospective users, as determined by detailed surveys of customers' preferences.

2. Assist a prospective customer in planning an installation. This type of design is usually made from sketchy specifications and can be quite approximate.

3. Estimate the physical characteristics and the cost of an apparatus so that the manufacturer may bid against competitors in the market place. When the specifications are exacting and when some very special characteristic is involved, a fairly detailed design may be required before the manufacturer can be sure of building the apparatus at reasonable cost.

4. Satisfy one rating of a line of standard units. Here the design problem differs from the custom-made apparatus in that many designs are made before an order is received in accordance with industry standards. Families of designs using duplicate components whenever possible and all having the same general appearance are usually found here.

5. Study the effect of using new materials, and of applying new technological developments.

6. Study the effect of new manufacturing tools or techniques.

7. Study the economics of changing the number or variety of "building block" components, in order to optimize practices.

8. Study the effect of changing the specifications of an entire product line, or the standards, on the cost and performance.

Principles

Engineers in an established product business with many years of available design knowledge do not often think of the principles which guide them in their design work. Nevertheless such principles exist; without them it would be impossible to design successfully. Engineers in a new product business are searching for principles, which often are not determined quickly or easily.

The principles of design are those factors, technical, practical, or esthetic, which guide designers in the setting up of design practices and in the making of design decisions. Some of these may apply only to a particular product; others may be quite general and may apply to broad families of products. The designer on the basis of his knowledge of electrical engineering fundamentals must understand the relationships between the design variables, such as flux density and current density, and the performance of the apparatus. He must be aware of the effect of variable loading on the efficiency of the apparatus: the effect of design geometry on the economics of material utilization. The designer must know the tools which exist in his factory, their capabilities and limits. He must be able to visualize the physical appearance of his apparatus before it is built. The customer must be continually kept in mind and his operating and maintenance problems adequately provided for. A sufficient number of safety factors must be built into the apparatus to prevent failure in service on overload or during abnormal conditions. The designer must be familiar with all industry standards which can affect the design of his particular apparatus and should display leadership in suggesting new standards which will benefit his customers and the industry in general. Especially for an established product business, such as transformers and rotating machines, the principles and rules of design are usually available to the designer in a "standard-practice" book, which lists the formulas and structures to be used, the standards to be met, the available tools in the factory, and other useful information. Standard practice is being continuously improved as a result of new technological developments, new standards, new ideas, and changing market requirements.

Regardless of the type of apparatus he is responsible for, the designer's frontiers are limited by the lack of sufficient technical knowledge, his inability to visualize simultaneously the interrelationships of all design variables, and the inadequacy of cost data. He is never quite sure to what extent a particular material should be stressed or what are the economics of changing a particular design feature. He is part scientist, part artist, continually balancing his technical knowledge with good judgment, making decision after decision, and calculation after calculation, as he moves along the design process.

Process

Once the principles of design are known and clearly understood by the designer, he may be given a set of specifications and told to prepare a design. The design process might be defined as the step-by-step attempt to satisfy the purpose of design by using the principles of design. The design process may take several forms. If the apparatus is simple, with few variables, it may be possible to construct a chart or nomogram which relates all the variables to the performance so that once the desired performance is known, values of the design variables may be immediately determined. Unfortunately, few products are so easily designed.

In the more normal case there are many design variables, some continuous, others discrete. The relationships between the variables and performance are usually nonlinear and the variables may sometimes be buried in complicated mathematical expressions such as transcendental functions, Fourier series, etc. There are usually more unknown independent variables than there are performance criteria so that any number of designs may be made to satisfy the specifications. The problem is, then, if time is available, not only to find one solution which will satisfy the specifications, but also to find the best solution. Hundreds of design rules may exist, all of which must be satisfied for proper performance and to ensure that the design being made is economical. Some of the components or building blocks to be used in the design may be completely fixed and only a few of them available. To synthesize such a complicated matrix of continuous and discrete variables into a rigorous, intelligible design process is a herculean task.

Fig. 1 shows schematically the process of design and the flow of information. The design process may be split into three parts: 1. The preliminary design, which is usually sufficient for a proposition, a design study, and the start of a requisition. 2. The final design, which is essentially the refinement of the preliminary design to the most minute and complete details. Its output is the final design engineering record and the drafting ond manufacuring instructions. 3. Checking of the final design, usually by another designer.

As shown in Fig. 1, both the preliminary and final designs draw heavily on the information contained in the standard-practice book and only seldom, in unusual cases, particular problems must be solved for which the information is not available. In this case the experts will be consulted. These are the more experienced designers, development engineers, and specialists in electrical, mechanical, and thermal problems. The standard-practice book embodies the best knowledge available in the design office and is being continuously revised, updated, and improved by the experts. In conclusion, with this system, during the design process the designer is very seldom called upon to solve complex mathematical equations, to analyze fields, to invert matrices, etc. In most cases these steps have already been taken for him by the experts, and the information is available in the standard-practice book in the form of tables, charts, nomograms, and predesigned structures and components. Thus, his job is to perform rather simple but lengthy calculations, usually with a slide rule, and to select, apply, and integrate the available information. In many ways, in addition to engineering calculations, he is doing engineering data-processing work. This point is important and will come up again later in discussing the type of computer to be used for design.

The time-honored design process for most apparatus is the cut-and-try method. The designer estimates values for all of the independent design variables, calculates the physical dimensions of all the components of the apparatus and then, for certain loading conditions, calculates its performance, such as losses, regulation, and temperature rises. These are compared with the desired performance. If the calculated values differ from those desired, the designer estimates in what direction and to what extent the independent variables should be changed, and repeats the calculations. This process is continued until the calculated performance is within acceptable limits of the desired performance. The length of time required to converge a design in this manner depends largely on the experience and ability of the designer. An inexperienced man will often discover to his chagrin that changes in values of the independent variables have resulted in a design which is further from desired performance than the previous design. The number of steps which the designer must take as he moves along the design process depends not only on the complexity of the apparatus but also how near he is to satisfying the required performance. Usually on the first couple of iterations the designer will make many approximations and avoid detail as much as possible. As the design "homes in" it is necessary to describe components in more detail and make more comprehensive and accurate performance calculations. This latter state may entail thousands of small steps such as selecting particular pieces of design information, integrating these with other pieces of information, calculating physical size, checking all design information to ensure that it is still adequate, and calculating performance. The designer may sometimes make sketches of circuits and components at this latter stage. When he is reasonably certain that the design is what he wants, he will write out the design output documents which contain all of the important information.

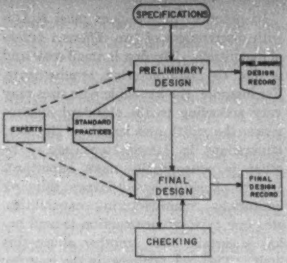


Fig. 1. Design process and flow of information

— normal flow
– – occasional flow

Most persons who have become genuinely interested in the design of electric apparatus have probably at some time in their careers been tempted by the idea of trying to synthesize manually the design process by methods other than cut-and-try. Multivariable curves, nomograms, and other short-cut techniques have been tried by the hundreds. Some of these are excellent approximations to a design and help the inexperienced designer to select the proper core size or punching dimensions as a start. Few of these are accurate or detailed enough to yield a complete design because of the limits to the number of variables which can be handled by 2-dimensional graph paper.

A third process of design which has been successfully used is the "automatic-convergence" method. This approach is neither cut-and-try nor pure synthesis, but it contains some of the features of both. This process entails the estimation of some of the independent variables and then a step-by-step synthesis of the design, usually one or two variables at a time. As the process proceeds, the design variables are continually improved by convergence techniques until they cease to change. When this occurs the design has been accomplished. The automatic-convergence process is very difficult to describe for any complicated apparatus since it requires that design judgment be built into the process. Once it has been described in an intelligible fashion, however, the design process can be carried out by a clerk rather than an engineer. In fact, an engineer would not wish to use this method of plodding along mechanically since he would immediately spot short cuts in a specific design to speed the process. Yet the automatic-convergence method, as will be seen, has proved itself invaluable in the application of digital computers to design.

EFFECT OF DISCRETENESS OF VARIABLES

In the foregoing it has been tacitly assumed that the design variables could be varied continuously in infinitely small increments by the designer within the predetermined limits. Obviously this is not always the case in practice, because of manufacturing requirements, available tools and dies, industry and company standards, and sometimes common usage (for instance, the use of fractions of an inch). As the modern electrical industry is becoming more standardized and more automated, this discreteness of the design variables becomes more and more important and thus warrants further analysis and discussion. Normally, some variables (for instance, flux density and current density) are continuously variable, others such as wire dimensions are discrete in smaller or larger steps. All combinations of continuous and discrete variables are theoretically possible and, in fact, great variety is found for different apparatus and for different manufacturers. To visualize better the importance of this problem, consider two extreme idealized cases: 1. the completely custom-built product, and 2. the product completely assembled from available standard components. In the first case of the custom-built product, such as a large power transformer, the designer has all the variables at his command without restrictions, except for the fundamental laws of science and certain upper and lower limits of manufacturability. His job to determine the sets of simultaneous values of these variables which, subject to the boundary conditions, will minimize or maximize the dependent variable which has been assumed as the "value criterion," for instance, weight, cost, over-all dimensions, or any combination of these and other variables. If all the design equations were linear, the problem would be akin to linear programming, conceptually simple, although the calculations might be quite involved and laborious.

However, since most design relationships are nonlinear, the problem is much more complicated and it is difficult to visualize without previous experience in the art. One simple method of visualization is the following.[11] The relationship of the dependent variable (say, total weight) versus the independent variables may be represented geometrically as a multidimensional surface in a multidimensional space. Assume, for illustration, that there are only two independent design variables. There will be, then, an ordinary surface, on which the lowest point within the boundaries must be sought. This problem is very similar to that of climbing the highest peak in a mountain range without a map. One easy way is the "method of steepest ascent," that is, to keep climbing always along the steepest possible path. Sooner or later a summit will be reached. This peak represents a local optimum (or suboptimum in the parlance of operations research[12]) but there is no guarantee that it is the highest summit of the range, the optimum *optimorum*. In fact, the highest summit may be much higher than the one reached and it may be far away.

Consider now the second case, wherein the product is to be built as an assembly of available standard components. (Obviously, the major design problem is the choice of the type and number of these components but, for the time being, these will be assumed given.) Now all variables can vary only according to a limited number of discrete steps, and all the combinations of these steps are not possible. The pleasant rolling hills of the surface now become a series of jagged ridges and peaks. The "method of steepest ascent" is of no value, since the climber is faced at every step by vertical walls. In this case the problem is to find the optimum combination of the possible values of the variables, and almost always it will be necessary to sacrifice progress along one variable, to achieve a larger gain along the others. It is also evident that, other things being equal, in this second case the total weight will be generally somewhat greater than in the first case. That is, a reduction in the number of possible steps is usually accompanied by an increase in weight, but not necessarily in cost, a fact to be discussed extensively later.

In practice, the designer must reckon with discreteness of two different orders of magnitude. The first is small-scale and is due to manufacturing requirements. For instance, the available wire sizes may vary according to the standard number series, the width and length of the steel laminations in 1-inch steps, etc. The other is conceptually more important. Due to physical characteristics, shipping limitations, manufacturing capabilities, etc., one type of construction is used below a certain limit, another above this limit. This, for instance, is the case for high-voltage windings in core-type transformers wherein two widely different constructions, continuous-disk and layer-wound, are used respectively below and above a certain BIL (basic impulse insulation level). This BIL is not absolutely fixed, but may vary somewhat, and there is a "grey zone" of overlap.

This phenomenon of discreteness is inherent in the design process, and to ignore it would be an inadmissible simplification of the problem. In fact, with the present trend of the electrical industry towards standardization, mass production, and automation, discreteness is becoming more and more important.

It is also important to note that different knowledges, skills, and types of personality are required of the designers for the two cases just considered. The designer of the custom-built product should be analytically minded and accurate. His job is to use the most refined and accurate methods of analysis so that the specifications may be met without any waste in material. The more accurate he is, the fewer will be the required "safety factors" and the more economical the design. The point of diminishing returns in his work is reached when the time and effort spent in refining a calculation is equal to the cost of the material saved. His skill is similar to that of a mathematician. The designer of the product assembled from standard components requires less analytical but more creative ability (in the artistic sense). His job is essentially to visualize the best opportunities in selecting the components and putting them together. His is a skill similar to that of the architect or the draftsman. Finally, both analytical and creative skills are required for the most difficult job in design: the establishment of the type, number, and characteristics of the standard components.

Impact of Digital Computer

Both analog[13] and digital computers have been used for the design of electric machines. It is safe to say that the digital computer is now far ahead of the analog in the broad design area. This is principally because of: 1. the incapacity of the analog computer to handle the discreteness of the design variables, and 2. the truly explosive progress of the digital computer in the last few years. Harder[14] states that there has been a ten-to-one increase per year in computing power, as measured by the speed, size of fast-access memory, flexibility, and reliability. Also, there is no evidence that this ratio of progress will diminish in the foreseeable future. There has been a parallel, and equally striking, reduction in the time and cost per computation. Harder states that a computation requiring 2 weeks on a desk calculator, and costing $300 only a few years ago, now costs 3 cents on a high-power computer, a reduction of 10,000 to 1. The ratio of reduction in time is even higher, with present commercial computing speeds as high as 40,000 additions per second. It should be noted, however, that these cost figures do not include the programming expense, which may be quite considerable in dollars and man-years.[15]

The characteristics of modern computers, the principles of operation, and the programming techniques are well known[14,16] and will not be repeated here; only the principal characteristics will be listed which make this new and powerful tool extremely well adapted and useful to electrical design: 1. high speed and low cost per calculation; 2. flexibility; 3. large memory, of which a considerable portion has extra-fast access; 4. high input and output speeds; 5. logical ability; 6. high speed in making logical decisions; 7. the possibility of achieving an unlimited number of loops in the program, by using the stored-program and address-modification techniques; 8. extremely high reliability and accuracy; and 9. spectacular and continued improvements in the hardware, as has been mentioned, and in the programming art.

The justification for applying computers to the design of electric machines arises from these characteristics and is threefold: computers, when properly applied, save time and money and, most important, make possible calculations which would be impossible by other means. This last factor, in turn, contributes to technological progress and to additional savings. The outstanding characteristics of the digital computer having been reviewed, the discussion will now turn to its application to electrical design and future developments in this field.

Vol.
Pt. 1
1958

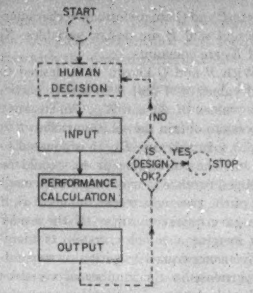

START
HUMAN DECISION
INPUT
NO
IS DESIGN OK?
YES
STOP
PERFORMANCE CALCULATION
OUTPUT

Fig. 2. Flow of information for design analysis method

— computer
– – human

Application of Digital Computers to Design

Design Analysis

In applying the computer to electrical design, the first method naturally coming to the mind of the designer is to use it as a glorified slide rule for performance calculations or design analysis. Considerable advantages over calculations by hand have been obtained by this method, especially when it is possible to introduce in the program more refined and accurate calculations than are used in the hand process. Essentially, in the analysis design method, the designer estimates the design variables and the computer calculates the performance of the apparatus. In most cases the calculated performance will not agree with the specifications, and the designer must modify the initial estimates. This 2-step process is repeated as many times as necessary to meet the specifications within the allowable tolerances. Thus the design loop has a human portion (the modification of the estimates) and a computer portion (the performance calculation), as shown in Fig. 2. This human intervention in the program has two major disadvantages. The first is common to both design by hand and by computer. Since the modification of the estimates depends upon the ability, judgment, and experience of the designer, there is no guarantee that in the successive loops the results of the calculation will come always closer to the specifications, and that the design is gradually improved. In fact, an inexperienced designer might conceivably waste much computing time and never achieve the desired results.

The second disadvantage, to use an electrical metaphor, is due to the mismatch of impedance between the designer and the computer. Medium-sized and large-sized computers presently used in design are between 100 and 10,000 times faster than the human mind. A considerable portion of the expended time is required for data input and output, rather than in performing calculations. Thus the efficiency of the design loop is very much lowered by its weakest link, the designer, and by the necessity of reading the new input and printing out the new output instead of keeping it within the memory. Furthermore, as has been stated, there is a marked trend towards faster and better computers. Yet, the capacity of the human mind will remain more or less the same and thus will become always a greater bottleneck in the design loop. The design method of analysis, however, has certain advantages[8] and may be the best in introducing the computer to a group of engineers, who may initially regard it with suspicion, as a potential threat rather than as a willing and docile helper. This approach also avoids the actual or imagined fear of centralizing all design knowledge within the small computer staff. From a psychological point of view, this may be quite valuable in the initial stage, since the individual design engineer has complete control over the making of decisions in the design process. Thus his prerogatives and responsibilities have not been encroached upon.

The next obvious step is the removal of the human element from the design loop. This requires the development of a new system, conceptually different from the hand system, since now the machine must obtain by itself without human intervention a design which meets all the specifications. Or, if this is not possible, it must point this out in the output. Thus the process is conceptually very simple (Fig. 3, contrasted to Fig. 2). Since all decision making is now entrusted to the machine, it is necessary to develop a foolproof, or almost foolproof, automatic procedure. For a smaller computer, for lack of storage capacity and speed it may not be possible to concentrate all decision making to the machine. One solution is to let the designer make initially the major decisions concerning alternate methods of construction for instance, disk versus layer winding; and to let the computer obtain a design with this type of structure which meets the specifications. Of course, this design may not meet the manufacturing or shipping restrictions, or be the best. Then, the designer must try out a new one. The next step is, of course, to have a complete design synthesis wherein, for valid specifications, the computer will always achieve a design valid from all points of view.

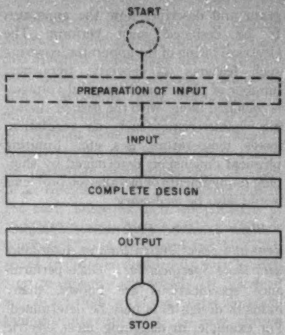

START
PREPARATION OF INPUT
INPUT
COMPLETE DESIGN
OUTPUT
STOP

Fig. 3. Flow of information for design synthesis method

— computer
– – human

Design Synthesis

A practical method of synthesis is shown in Fig. 4. Although this method was first developed by the authors for the transformer design problem, the principles have been extended to other types of products. It is felt that these ideas are general enough to be applicable to all electric apparatus. Each of the blocks will be described in detail in the following, with particular emphasis on the over-all problem of design synthesis.

Block 1. Performance Specifications: This is the input to the computer pro-

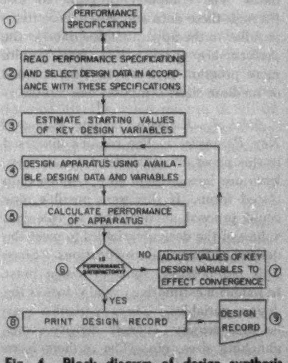

①
PERFORMANCE SPECIFICATIONS
②
READ PERFORMANCE SPECIFICATIONS AND SELECT DESIGN DATA IN ACCORDANCE WITH THESE SPECIFICATIONS
③
ESTIMATE STARTING VALUES OF KEY DESIGN VARIABLES
④
DESIGN APPARATUS USING AVAILABLE DESIGN DATA AND VARIABLES
⑤
CALCULATE PERFORMANCE OF APPARATUS
⑥
IS PERFORMANCE SATISFACTORY?
NO
ADJUST VALUES OF KEY DESIGN VARIABLES TO EFFECT CONVERGENCE
⑦
YES
⑧
PRINT DESIGN RECORD
DESIGN RECORD
⑨

Fig. 4. Block diagram of design synthesis method

gram and describes how the apparatus to be designed must perform. The electrical rating of the apparatus, covering such features as capacity, frequency, number of phases, and terminal voltages, is included along with the desired operating characteristics such as regulation, losses, temperature rises, etc. Limiting physical dimensions determined by shipping requirements or installation problems would also be necessary.

Block 2. Read Performance Specifications and Select Design Data in Accordance with these Specifications: Once performance specifications are known, many items of design data may be determined. For example, in designing most electric apparatus, when the BIL is known the minimum electrical clearances between various points in the apparatus are fixed. The types of components to be used and the data to describe them may depend on whether the apparatus is to be single-phase or 3-phase. Some design practices may change with the capacity of the unit and recognition may be made of this. Since physical size is unknown at this point, any design data which depend on it should be estimated from available information. These data will be revised as the design proceeds and as information concerning physical size becomes available. The amount of information which must be available at this stage in the design process depends largely on the complexity of the apparatus involved and the scope of the computer program. A narrow program applying to one type of construction for a simple apparatus might require very few items of design data. A broad program applying to several types of construction for a complex apparatus could need thousands of items of design data. The computer program to sort through these data and select those items peculiar to the apparatus described in the performance specifications may require more program instructions than a small or medium-sized computer can handle.

Block 3. Estimate Starting Values of Key Parameters: For reasons discussed in this paper, most practical design problems are incapable of direct solution in closed form. If this were possible, one could proceed at this time to calculate values of the design variables to meet the desired performance specifications. Since it is not possible, some provision must be made for estimating initial values of the design variables. These variables may be flux density, current density, core-punching sizes, etc. The authors have found that usually it is not necessary to be very accurate at this point because if the remainder of the program is properly constructed the design will automatically converge to the desired design regardless of the initial values.[2] The difference in machine running time between a very accurate start and a very inaccurate one is only 5 to 10 seconds on the IBM *705* computer, for a preliminary transformer design requiring 1 to 2 minutes in total.

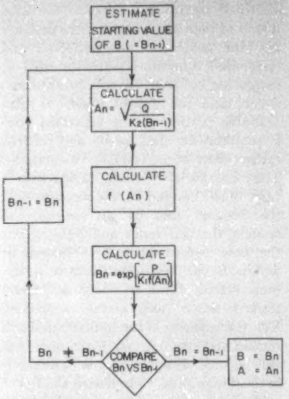


Fig. 5. Block diagram of iterative process

Block 4. Design Apparatus Using Available Design Data and Variables: This is the heart of any synthesized design program. It requires the greatest skill to describe and will probably use the largest number of computer instructions. The form that it takes will depend on the type of apparatus being designed and the degree of design completeness desired. The order of events within this block is quite critical in effecting convergence. The person who constructs this segment of the program must not only be very familiar with the apparatus and the design problems peculiar to it but he must also be extremely creative since little precedent will exist for the necessary manipulations.

During this phase of the design process, new values of the design variables will be generated which will be better than those available at the beginning. The authors have found that iterative techniques are very useful at this stage, particularly when the performance relationships involve nonlinear functions. For example, suppose that the following relationships exist:

$$P = K_1[f(A)] \ln B$$

$$Q = K_2 A^2 B$$

where P and Q are performance characteristics, A and B are design variables, K_1 and K_2 are constants.

With P and Q known, it is desired to find values of A and B to satisfy particular values of K_1 and K_2. An iterative process to obtain the solution is shown in Fig. 5. In this case $B_n - 1$ is equated to B_n but the converging process could be changed to some form of weighted averaging of the two most recent values of B, if this gave quicker results. If the reader can imagine a complex system of many performance equations with a corresponding increase in the number of variables and with dozens of limitations upon the system, he has some picture of the typical design process. The authors have discovered that it is usually possible to describe such a complicated system by a series of relatively simpler systems. These new systems are simpler in that only two or three variables are allowed to vary at one time and the remaining variables are held fixed at that time. The results of this process are carried over to the next system and there are many iterative loops within other iterative loops. The order in which the smaller systems are processed depends on the relationships between variables. As with any iterative procedure, to effect convergence the most sensitive variables must be dealt with first, the next sensitive second, and the least sensitive last. As values of the variables are improved, better information about the physical size of the components becomes known. Any design data which depend on physical size must be continually checked for validity and changed if necessary.

Whenever a question arises as to which is the best way to proceed, for example, which is the best type of component to use, how should the components be arranged, what balance of active materials should be used, the program must contain all the criteria for making the decision. The judgment which has been traditionally exercised by the designer must be built into the computer program so that the machine will always make the same decision as the man, under the same set of circumstances. The subsequent analysis of various design methods and the discarding of all but one which will be used in the computer program is one of the most rewarding aspects of this type of work.

At the completion of stage 4, values of all the design data and variables will be known. There is no guarantee that this is the proper design, however, since several estimates will still be factored into the design. The only way to prove the correctness of the design is to calculate the complete performance and compare this with the desired performance.

Block 5. Calculate Performance of Apparatus: During this stage the computer will perform the design analysis and calculate each of the performance characteristics of the apparatus. Some of these calculations may be simple; others may be mathematically very complex. This is a relatively easy stage to describe by block diagram and to program, because of its straightforwardness and because it ordinarily does not involve judgment.

Block 6. Is the Performance Satisfactory? Each of the calculated performance characteristics must be compared with the desired values to see if it is satisfactory. Criteria for acceptability must be set up. For example, it might be desirable for one characteristic always to be below the desired value. Another characteristic might have to be with ±1% of the desired value. If all the calculated values are within acceptability limits the design can be said to be satisfactory. If one or more of the calculated values are outside of acceptability limits the design is not satisfactory and at least one more iteration of stages 4 and 5 is necessary.

Block 7. Adjust Values of Key Design Variables to Effect Convergence: Because the design is not satisfactory, values of the design variables are not correct and must be improved. Knowledge of the interrelationships between variables is most important for determining how the variables should be changed at this point. Possibly the most recent values developed in stage 4 should be used. This is usually the case except where it induces oscillations between values used to start stage 4 and values available at the completion of stage 4. For this oscillating case a system of weighted averages will usually suffice and the weighting factors may be used to speed up convergence. Rather than spend much time worrying about this problem, the authors have discovered that it is better to complete the remainder of the program and use the program to study the behavior of the design process in order to determine the best solution.

Blocks 8 and 9. Print Design Record: After the design is satisfactory, a complete description of the physical characteristics of the apparatus, along with the calculated performance characteristics, should be printed. Any problems which the machine encountered during the design process should also be printed out for the convenience of the design engineer who will evaluate the design.

One advantage of the synthesis method described is that it allows the apparatus to be designed in whatever detail is desired. If a very rough design is adequate, stages 2 and 4 may be made small and simple. If the program is to be used for generating instructions on how to construct a complex apparatus, stages 2 and 4 could become so large that they would require a large-scale computer. The authors have discovered that there are advantages to thinking of the total design process as having two distinct phases, preliminary design and final design; see Fig. 1. The preliminary-design phase is primarily concerned with the design of the major components of the apparatus and omits any detail which does not significantly affect the weight, physical size, performance, and cost of the apparatus. The final-design phase supplements the preliminary design by generating all the additional detail required to describe the apparatus for manufacture. Some advantages of this concept are:

1. In most product businesses, many designs are required for apparatus which will not be built, for example, to answer customer inquiries and for design studies. A preliminary design is adequate for such purposes.
2. A final design will require more computer programming effort than preliminary design. Also it will not be run as frequently.
3. A final design will take much longer to run on a computer and therefore should not be run unless the detailed information is necessary.

It has been the authors' experience that the preliminary-design phase is the "workhorse" of an engineering organization which designs custom-built apparatus.

Conceptually there is a major break between preparing a program for design analysis and one for design synthesis. This is due to two basic differences. The first is that performance calculations are essentially mathematical with few logical decisions, and are most appropriately done on a computer of the scientific type. Design synthesis on the other hand, as we have seen, entails a tremendous amount of selection and processing of data and logical decisions, which may require more time than the actual performance calculations. A computer of the "business" type is probably best suited for this application. In fact, optimum utilization of the computer memory and speed may be obtained by using a machine with variable word length, such as the IBM *705*. This machine has high speed in reading, writing, and making logical decisions, but relatively slower speeds in making additions, subtractions, multiplications, and divisions, which may amount to only 30% of the total instructions. Usually the 10- or 12-decimal digit word length of scientific computers is excessive for design applications which, as is well known, usually require slide-rule accuracy except for particular subroutines such as reactance calculations.[17] Recognizing this in the early stage of their apppication of the CPC to transformer design, a 5-digit plugboard[18] was developed for the CPC, which in practice doubled its very limited storage capacity and speed. It is indeed significant that two of the largest electrical manufacturers in the U. S. have chosen an IBM *705* (a "business" machine) rather than an IBM *704* (a "scientific" machine) for transformer design synthesis.

The other basic difference between programming design analysis and synthesis is the following. The rules and formulas for performance calculations are readily available from the fundamental laws of electricity, textbooks, standard practice, and test results. These may be simple or complex but they are available. In contrast, the rules for making decisions in design synthesis, for selecting components, for modifying the initial estimates, etc., are usually not written down, but are hidden in the mind of the designer. In many cases these decisions will differ according to the ability, experience, training, and personality of the designer and it may be difficult to reconcile the various experts and "pin them down" to a unique set of rules. In the hand method this may not be a serious drawback and it permits the injection of the designer's individual characteristic into the product. As long as no clean-cut perference can be established for the choice of one alternative, it may be wise to avoid making a premature decision. With the advent of the computer, this avoidance of decision making must come to an end. It is necessary therefore to write down and describe to the most minute detail all these unwritten rules and practices, and choose one, and only one, of the possible alternatives. This may require tact, perseverance, and time in a group of engineers who have worked for many years, as described previously. In making these choices, it is only natural that the criterion is the "best" alternative in every case. This leads directly to the subject of optimization, to be discussed later. For the time being, note that there may be many different opinions as to what is the best alternative, and that in many cases there is simply not enough knowledge available for making a decision. Yet a decision must be made in order to write the program. It may be felt that this is a serious drawback, from an economic standpoint, since possible better alternatives will not be considered, and from a psychological one, since the individual designer may feel that he cannot use fully his judgment and experience. In some cases, many of the possible savings may be lost, if the designer modifies the computer-produced design by hand, to improve it according to his judgment or avoid using alternatives which he does not like.

Fully realizing these factors, and their incapacity of being absolutely sure of having the best alternative in the computer, the authors in their *705* programs for transformer design have adopted the policy that the designer may always override the program in many important fields. The program works this way: if only the specifications are furnished to the computer, it will produce a design according to the program, which embodies the "best" alternatives, as obtained from the best available judgment and experience in the design groups. By adding additional input, the designer may specifically override the computer program and obtain alternative designs, which he may compare, if he wishes, with the routine design. This flexibility is extremely valuable in studying new ideas, proposed new developments, or special customer requirements. It maintains an always "alive" design program and keeps the individual designers happy.

It is seen that, the design synthesis method leads, sooner or later, to an inescapable consideration: design optimization. This is a fascinating subject and herein still lies the tremendous future potentialities of this powerful new tool, the digital computer which, in the authors' opinion, has been used to only a minor degree of sophistication.

DESIGN OPTIMIZATION

Design optimization is the ultimate aim of every designer. It is the authors' conviction that true optimization will be possible only through the modern, high-speed digital computer. In considering the broad problem, one is faced immediately with that bugaboo of operations research: suboptimization. Why this is so is obvious: one person cannot optimize one variable and the next person optimize the others. This would happen in the example, given earlier, of the surface of hills and peaks, by selecting a fixed value of one variable and climbing to the highest point of that cross section. There is no guarantee that it is the highest summit, or even a secondary one. Thus, optimization is a complex subject for various reasons. First of all, at least theoretically, it cannot be accomplished by separation of the variables: it is an all-or-nothing process. Second, a "theory of value" must be establihsed; that is, a decision as to what to optimize. In many cases, including design, more time, money, and effort are wasted in trying to optimize the wrong variable, than in entirely omitting optimization. The "value" may be weight, dimensions, cost, quality, customer satisfaction; the balanced best interest of the manufacturer, the customer, and the public; or a combination of these quantities, which are progressively more difficult to define and determine mathematically. The point is reached where it is impossible to write a computer program for such elusive criteria. The over-all decision is best left to the judgment and experience of the manager who is responsible. This, of course, is the general approach of operations research which furnishes the pertinent data to the executive and allows him to make the final decision. Furthermore, optimization is strictly a relative criterion. A given design is only optimum subject to the imposed boundary conditions and to the criterion chosen as "value." Thus the design with minimum weight is seldom the design of minimum cost to the manufacturer. In turn, the design with minimum cost to the manufacturer is not always the design with minimum capitalized cost to the user, and the optimum design taking into account both the cost of the manufacturer and the user may not be optimum from the standpoint of the over-all economy, etc. Here again is the bugaboo of suboptimization.

To see how optimization is complicated and laborious, even for a large computer, consider the idealized case, described in the foregoing, of a product made exclusively of standard components. Assume that the product must meet the following three characteristics specified by the customer: 1. reactance, 2. current, and 3. system voltage. Assume that the reactance values must be met within ±1% and the current values within ±2.5%. The range of rating usually requested by the customers varies from 1 to 200 in reactance and from 1 to 36 in current, and there are nine standard values of system voltages. It is easy to see that the total number of designs to be made would be

$$\frac{\ln 200}{\ln 1.02}\times\frac{\ln 36}{\ln 1.05}\times 9=270\times 75\times 9$$
$$=185{,}000 \text{ designs}$$

At present, it is not known a priori how to optimize a design. Since there are seven independent design variables in all, many different combinations of these variables, say, 10 or 12 should be tried for each design by the synthesis method. Even after performing about 2 million designs, which may require 30 seconds each on a large computer (costing $500,000 per year), i.e., about 8 computer-years at 40 hours a week, or 4 million dollars, it is still uncertain as to whether the optimum has been achieved. Obviously the answer is to perform each design as it is required, and to try as many different variations as long as there is computer time available and as long as the expected amount to be saved is equal to the cost of making an additional design on the computer.

Another way to look at this problem is, so to speak, to turn it upside down. As has been mentioned, there are seven independent design variables. For each one are known the minimum and maximum values and the number of finite intermediate values. For instance, the value of one design variable usually varies between 4 and 20 and must be even, so that there are only nine possible values in all. Assume that there are ten possible values, on the average, of each design variable. Thus 10^7 or 10 million combinations in all must be tried. However, these are strictly performance calculations on the computer and since they are done according to a mass-production technique they require only one hundredth of the time of an individual design by the synthesis method, or 0.3 second each. Thus the total time required would be about 800 hours or about $200,000, which is not an unreasonable figure for this amount of work. Now that all possible combinations of the design variables are available, it is certain that they contain all 185,000 combinations of the customer specifications mentioned, if these specifications can be met at all with the available components. The next step is simply data processing and may even be done on conventional punched-card equipment. The 10 million designs are sorted in 185,000 packs, each corresponding to one set of customer specifications. Each pack, which contains, on the average, 54 alternatives, is sorted according to the computed cost, if this is the chosen criterion. The 53 more expensive designs are discarded and the optimum is retained. Now, it is absolutely certain that the optimum has been obtained, i.e., the optimum for the boundary conditions chosen. But the choice of the number of these variables may be questioned, for instance, the number of standard wire sizes, and so on, ad infinitum. This example, which was actually tried out by one of the authors, shows conclusively that: 1. optimization is a relative concept and, most important, 2. it will become possible only through the largest and fastest computers now available or planned.

In the practical realization of their transformer design program, which is very much more complicated than the program just described, the authors chose a realistic and down-to-earth approach towards optimization, and determined on the following procedure: 1. to try to achieve the optimum within the present status of knowledge of design; 2. to strive continuously to improve the optimum by developing new knowledge and to use the computer extensively for this; and 3. to utilize the store of experience in the mind of the individual designer by making it possible to override the computer program, as discussed in the foregoing.

Using manual design techniques, the designer is seldom able to investigate many alternative constructions to determine the best. Therefore it is extremely important for him to develop rules of thumb to guide him toward on optimum design. The authors have found that these same or similar rules, when built into a computer program, greatly aid the computer to reach an optimum design. Several examples taken from the transformer design problem will help to illustrate this.

Through the years, studies have been made on the optimum "form" of a transformer. Since the designer has almost complete control of whether a given transformer will be short and squat or tall and slim, this is extremely important. Based on existing rules of thumb a relationship between optimum transformer height and kva was derived and built into the program. Essentially, it stated that transformer height should increase with kva up to a maximum allowable shipping height. Above this point, maximum possible height gives an optimum form. The selection of the most economical type of winding to use for a particular design has traditionally depended on the axial turns per inch of core window and the total cross section of conductor desired. To verify these relationships, an extensive study was made on the CPC wherein several types of windings were designed to accomplish the same objectives. By relating costs back to some of the design variables, it was possible to set up rules concerning the selection of winding types and to build these rules into the computer program. Also, since it can be shown that optimum copper utilization is reached when the winding temperature rises of each of the windings are equated, this rule is built into the computer program. The combination of these and several other rules results in a design very close to optimum.

STANDARDIZATION

Consider again the first idealized case discussed previously, the custom-designed and custom-built product, with practically no repetition of ratings over the years. It has been seen that the point of diminishing returns on refining the calculations is met when the cost of such additional calculations becomes equal to the expected savings in material. Since the digital computer reduces drastically the cost of calculations, it would appear that considerable savings would be obtained by carrying all designs to the maximum detail and variability, eliminating any discreteness in the design variables. However, this concept is true only for the ideal case and it neglects manufacturability. In practice, electric machines are all more or less standardized and the practical application of the foregoing concept to its ultimate would represent a typical example of suboptimization and would run directly against one of the most marked and beneficial trends of the electrical manufacturing industry: standardization. This leads to a discussion of standardization and this can be done advantageously only by keeping in mind the optimization concept discussed in the previous section.

Standardization, both industry-wide and within a company, department, or plant, has been and is one of the most powerful tools in improving quality, keeping down the total cost and extending the benefits of mass production to wider and wider markets. Basically, its purposes are: 1. to reduce the total cost to the national economy of a given product line, and 2. to make possible increased production per office and factory worker. If it is considered that the requirements of the electrical industry double every 10 years, and that the available workers will increase only a few per cent in the same period in this country, the importance of the second factor may be visualized. Furthermore, no automation can exist without standardization, which in electrical design can take several forms: standard ratings, standard outside dimensions, standard sizes and shapes of major components, standard dimensions of minor components, and standard steps in one or more dimensions of otherwise variable components. These will be designated henceforth as "discrete parameters." The following discussion will be sufficiently general to encompass all these cases.

Here the goal of standardization will be assumed to be the minimization of the manufacturing cost of the product for a given series of specifications and industry or company standards. It is obvious that in the general case variability is required in the design variables in order to: 1. meet the variabilities of the customer specifications and other market requirements, 2. economize material, and 3. fit together the various pieces of the "jigsaw puzzle" as represented by the standard components. The selection of the type and number of discrete parameters is primarily a matter of optimization and, as has been seen, this is not a well-defined area. As has also been seen, the problem is compounded by the fact that all these discrete parameters are interrelated by complex nonlinear relationships, so that the design of one may have a major effect on several others. Thus, the separate optimization of each component does not guarantee an optimized product, and consequently all discrete parameters should be designed together within the framework of an integrated system. It is felt, however, that some general concepts may be developed as an aid in the solution of this problem.

In trying to minimize the total cost, all its major portions must be considered separately; material; drafting expense; engineering expense; manufacturing and clerical expense; and computer costs. It must also be realized that the problem is often seen differently by the design engineer, the design draftsman, and the manufacturing engineer, since each group may overemphasize the importance of the contributions of the items with which he is most familiar.

The design engineer is interested in maximum flexibility and in the best utilization of material which will meet the customer's specifications. As has been seen, this is obtained by having available a very large number of discrete parameters and the effect of the computer is to shift the minimum cost point toward an even higher number of these parameters, since it allows more refined and accurate calculations. Fig. 6(A) shows qualitatively the basic relationship between material cost and degree of standardization, which is inversely proportional to the number of discrete parameters, i.e., the greater the degree of standardization, the fewer the number of discrete parameters.

The drafting and engineering expenses

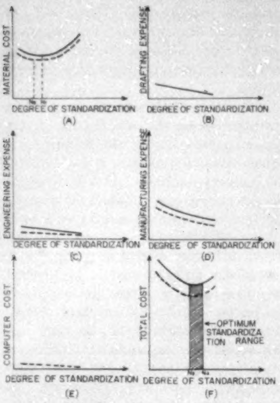


Fig. 6. Theoretical effect of degree of standardization on cost in dollars

— by hand
- - by computer

are basically proportional to the number of discrete parameters, but the slope of the drafting cost curve, Fig. 6(B), is steeper than the engineering cost curve; see Fig. 6(C). In fact, as the number of components, arrangements and discrete dimensions is increased, more layouts and detail drawings are required and more man-hours of creative effort will be expended in engineering and drafting. The effect of the computer on engineering expense is to reduce the slope of the curve, as shown in Fig. 6(C).

Manufacturing expense in general decreases with the increasing degree of standardization; see Fig. 6(D). It has also been claimed that a reduction of this expense will be obtained through computers. Finally, in applying the computer, the following must be considered: the cost of programming, preparation of data in a form suitable for computer processing (punched cards, magnetic tape, etc.), and capitalized running costs. This cost varies linearly with the number of discrete parameters, Fig. 6(E), but the slope is quite small, since the devising of a general program, suitable for a variable number of components is the overwhelming portion of the cost and because, in large machines, the running costs are hardly affected by the increased number of discrete parameters available on tape. Furthermore, if the discrete steps are properly chosen, they can be represented by algebraic equations, and the total computer costs become independent of the number of possible steps in the design variables.

The optimum degree of standardization is obtained when total costs are minimized. To do this, all five cost components of Fig. 6(A) through (E) must be added together, and each contribution weighed by an appropriate factor. The relative value of these factors will depend, just as the shape of each curve, on the particular product and manufacturer considered. For purposes of illustration, assume that material amounts to 40%, manufacturing expense to 40%, engineering, drafting (and computer, if any), expense to 20% of the total cost, assumed to be equal to 100%. Obviously the total cost after introduction of the computer will be lower than before, if the application is worth it, but the same relative proportions may be assumed. Then, the curves of total cost versus the number of discrete parameters will take the general shape of Fig. 6(F) with minimums corresponding to about the same value N_3 of discrete parameters, much less than N_1 and N_2 of Fig. 6(A). It is also noted that the curve of Fig. 6(F) is quite flat around the minimum, as in most similar economic studies, and that therefore there is a certain latitude in the choice of the optimum number of discrete parameters.

This is indeed fortunate, because so far it has been assumed that the problem is static rather than dynamic. As mentioned previously, electric power is doubling every 10 years with correspondingly increased production, and there is a definite rapid trend towards factory automation. This will increase the slope of the manufacturing expense curve and will thus shift the optimum number N_3 more and more toward the right. Since the installation and full utilization of a large computer and any standardization program involves several years of planning and preparation, the selection of the optimum value must be based on the foreseeable future rather than on "present" data, which are essentially past history. Thus, it appears logical to choose the optimum number of discrete parameters in the standardization range N_3 to N_4, as indicated in Fig. 6(F). In this manner will be avoided both the danger of overconservatism, which will lead to the maintenance of outdated custom-built products in this age of mass production and automation, and the excessive enthusiasm of overshooting the mark and "standardizing for standardization's sake" without considering the whole economic picture within the actual requirements of the market.

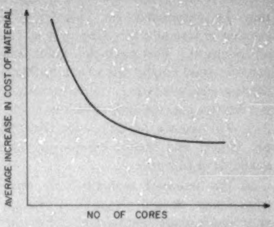


Fig. 7. Effect of number of core sizes on material cost

As an example, these standardization principles were used to analyze the optimum number of core sizes required in the power transformer business. Traditionally it had been felt that it was desirable to have a very large number of core sizes, since this allowed the designer to minimize the amount of active material, magnetic steel and copper, going into the transformer. The relationship between material costs and number of core sizes is shown in Fig. 7. Since most of the internal components of a transformer depend on the physical size of the core, a large number of core sizes results in more drawings, more factory paper, and higher manufacturing costs. An analysis of all factors indicates that it is theoretically possible to reduce the number of core sizes considerably. This requires more material in the product, but nonetheless reduces its total cost.

Finally, the computer, standardization, and automation all contribute to a splitting of the products marketed into two classes: the "standards" which may be designed on the computer and built from a majority of standard components, and the "specials" which must still be designed, for the major part, by the traditional hand methods and built in a job shop. As time passes, the split between these categories and the differences in cost are becoming more and more marked. This is bound to feed back and reflect itself on the market, for instance in price differentials, shorter deliveries, improved service and maintenance, etc. These advantages to the customer will create a larger demand for the standard product, compared to the special. The increased production in the standard area will sharpen more and more the advantages of the standards by allowing increased automation. Thus market and production will influence each other in a beneficial circle steadily progressing towards increased standardization and a better product.

Evaluation

A digital computer represents both high investment and a considerable annual expense. The investment is measured by the man-years required to prepare the design programs and by the cost of the physical installation. The yearly expense is measured by the rental costs, salaries of operating personnel, supplies, and overhead. For one shift, typical annual expenses may be of the order of $100,000 for a medium-sized computer and $600,000 for a large-sized computer. The initial investment may be of the same order of magnitude or larger.

Therefore it is very important to evaluate the contribution of a digital computer to design, and decide whether these high costs are economically justified. Evaluation of the contribution of the computer cannot be restricted exclusively to dollars and cents. This is only one aspect, albeit a very important one, of the entire picture. It appears that a particular computer application to design of electric apparatus' may be justified for one or more of the following reasons.

First, the computer saves money. The most obvious justification is that the cost in dollars to do the job by hand is so much greater than the cost of exactly the same job by computer, as in the classic data-processing application of payroll. This however, is only part of the picture for engineering applications, in that the computer design methods are usually superior to the hand methods, and therefore not directly comparable. In fact, engineers tend to use the computer not to repeat narrowly and unimaginatively what is being done by hand, but rather to obtain the best possible design, with regard to quality, accuracy, speed, and total cost. Actually, as has been shown, the savings obtainable through design optimization are considerably greater than the savings obtained through the reduction of the cost per operation.

Second, the computer saves time. A job requiring weeks by hand may be done in hours or minutes on the computer. This may be very important in giving faster service to the customer, and it is a real asset with respect to competition. Ultimately, it contributes to increased customer good will, better service to the public, higher prestige for the company, and a greater share of the market.

Third, the computer improves the quality of the product because it is usually more accurate than hand processing, and because errors are practically eliminated, once the program has been thoroughly checked out. It may be assumed that the quality of designs being produced by hand varies according to a Gaussian distribution. The effects of the computer are: 1. to raise the mean, since the programs embody the knowledge and ability of the most able and experienced designers rather than that of the average designer; 2. to eliminate almost completely the low-quality tail of the distribution curve for the reasons already stated; and 3. in some cases, to shorten a little the high-quality tail, since able individual designers may create original solutions best suited to a particularly complex job, whereas only the general solution is applied by the computer to these unusual cases.

Fourth, the computer allows the doing of things which cannot be done without it. This is probably the most important contribution and also perhaps the most difficult to grasp in its ultimate consequences. Engineers well know that there are problems, some extremely important, which simply cannot be solved by hand. A typical example is the optimization problem of a product assembled from standard components, mentioned earlier. It requires the computation, analysis, and evaluation of millions of alternate designs, an absolutely impossible task by hand.

Fifth, the computer frees the design engineer from daily routine calculations to truly creative effort, which may be devoted to the improvement of the product and to better service to the customer. It also alleviates the effect of the present shortage of engineers and technical personnel. By making design work less routine and more interesting, more students will be attracted to this field.

Conclusions

The application of digital computers to the design of electric apparatus is at present an accomplished fact. While this application is still young, it has achieved to date important results, technically and economically. The results to be achieved in the future appear even more important, particularly in the fields of design optimization and standardization. The present and future benefits of applying digital computers to the design of electric apparatus may be summarized in terms of lower costs, improved quality, faster and better service, more interesting and rewarding jobs, increased productivity, and more creative effort on the part of the engineering personnel.

References

1. How Digital Computers Aid Transformer Designers, **S. B. Williams, P. A. Abetti, E. F. Magnusson.** *General Electric Review*, Schenectady, N. Y., vol. 58, May 1955, pp. 24–25.

2. Application of Digital Computers to Transformer Design, **S. B. Williams, P. A. Abetti, E. F. Magnusson.** *AIEE Transactions*, vol. 75, pt. III, Aug. 1956, pp. 728–35.

3. Digital Computers as an Aid in Electric-Machine Design, **R. M. Saunders.** *Ibid.*, vol. 73, pt. I, May 1954, pp. 189–92.

4. Application of a Digital Computer to the Design of Power Transformers to Specification, **C. L. Moore, W. T. Duboc, P. A. Zaphyr.** *Ibid.*, vol. 75, pt. I, May 1956, pp. 134–38.

5. Product Design by Digital Computers, **M. Middleton, Jr.** *Westinghouse Engineer*, Pittsburgh, Pa., vol. 16, 1956, pp. 39–43.

6. Induction Machinery Design Being Revolutionized by the Digital Computer, **C. G. Veinott.** *AIEE Transactions*, vol. 75, pt. III, 1956 (Feb. 1957 section), pp. 1509–17.

7. Digital Computer Speeds Transformer Design, **C. P. Kappeler.** *Allis-Chalmers Electrical Review*, Milwaukee, Wis., First Quarter, 1957, pp. 22–25.

8. Electronic Computers—Engineering Timesavers, **F. J. Maginiss.** *General Electric Review*, Schenectady, N. Y., vol. 59, May–July 1956, pp. 14–17.

9. Management Appraisal of a Univac System in Operation, **S. Merriman.** "Electronics in Action," *Special Report No. 22*, American Management Association, Inc., New York, N. Y., 1957, pp. 866–96.

10. Transformer Design with Digital Computers, **J. V. Oldfield, D. McDonald, M. W. Humphrey Davis.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 103, pt. B, Supplement no. 1, 1956, pp. 54–58.

11. Let's Really Put the Computer to Work, **D. D. McCracken.** *General Electric Review*, Schenectady, N. Y., vol. 60, July 1957, pp. 13–15.

12. Suboptimization in Operations Problems, **C. Hitch, R. McKean.** "Operations Research for Management" (book), **J. F. McCloskey, F. N. Trefethen.** The Johns Hopkins Press, Baltimore, Md., 1954, pp. 168–86.

13. MONECA—A New Network Calculator for Motor Performance Calculations, **Cyril G. Veinott.** *AIEE Transactions*, vol. 71, pt. III, 1952, pp 231–39.

14. The Computing Revolution, **E. L. Harder.** *Electrical Engineering*, vol. 76, June 1957, pp. 476–81; July 1957, pp. 586–90

15. The Practice and the Economics of Applying Digital Computers to Engineering Problems, **P. A. Abetti, S. B. Williams.** *AIEE Transactions*, vol. 77, pt. I, July 1958, pp. 331–42.

16. Digital Computer Programming (book), **D. D. McCracken.** John Wiley & Sons, Inc., New York, N. Y., 1957.

17. Transformer Reactance Calculations with Digital Computers, **L. Rabins.** *AIEE Transactions*, vol. 75, pt. I, July 1956, pp. 261–67.

18. A New Five-Digit Plugboard System for Card-Programmed Calculator, **S. B. Williams, Miss N. M. Noonan.** *Ibid.*, Mar. 1956, pp. 67–71.

◆

Discussion

H. R. J. Grosch (International Business Machines Corporation, New York, N. Y.): The authors made a rather surprising choice when they adopted the IBM *705* for design synthesis. The variable word length feature which is so valuable in accounting work, where data fields may range from two or three characters to hundreds, is not utilized very effectively in technical work, nor is the manipulation of alphabetic information. The much greater arithmetic speeds of the "scientific" machines, and the availability of very much larger random-access memories for even more sophisticated design syntheses which the authors will undoubtedly attempt in the near future, point to something like the new IBM *709*.

In my experience, 30% is an unusually large fraction of instructions to be arithmetic as distinguished from logical. Only a very few programs, notably those like nuclear reactor designs and flutter calculations which involve large matrices, would require a higher proportion. In business applications which involve much rearrangement of data or maintenance of large tape files, arithmetic instructions may constitute only 2% or 3% of the programs.

The General Electric Company's success at Pittsfield and at Hanford in using *705*-type equipment for technical calculations, and Lockheed's success in using *704*-type equipment for business data processing (production control), point up emphatically the flexibility of the big digital computer. Almost any machine can contribute powerfully to almost any class of large-scale data handling. The key is, as always, able personnel and imaginative leadership.

C. G. Veinott (Reliance Electric and Engineering Company, Cleveland, Ohio): This paper was most interesting to me. The authors seem to have started about the same time that I did in putting a digital computer to work on electric machinery design. In the summer of 1953, I was responsible for having had an induction-motor design procedure programmed on a CPC.[1] Later, we followed this by programming a d-c design calculation on the CPC, also d-c armature heating calculations. This experience led us to the decision to acquire a digital computer for our own use. At that time, it was clear that the CPC was already outmoded and we turned to a stored-program computer, the Alwac III. Why we chose this particular computer and how we were then using it for induction machinery design calculations were reported in reference 6 of the paper.

The paper is an excellent discussion of the problems that face one who attempts to apply the digital computer to the practical design of electric machinery. These problems are quite distinct and different from many of the sophisticated techniques used and reported in the mathematical journals. One reason, as the authors point out, is that so many variables are discrete that many optimization techniques are not applicable.

As reported (reference 6 of the paper), we used the design-analysis method first. The first calculation procedure that we programmed on the CPC was that which was then being used by hand, and which required approximately 2 to 3 hours for an engineer or technician. In preparation for a stored-program computer, we developed new procedures and formulas to get a more accurate and more universal method. Some 10 to 12 man-months were expended in the development of the new calculation procedure. Calculations by this procedure required some 8 to 10 hours to perform by hand. This procedure was first programmed on a Datatron, then some refinements were added and it was later programmed on our Alwac. This emphasizes the point made by the authors that with a digital computer it is practical to use more elaborate and more accurate calculating formulas. However, unlike the authors, we did not find these formulas readily available from textbooks.

We used the design analysis method for starting for at least three reasons:

1. If permitted a mass tryout of our new calculating method.

2. It introduced engineers to the computer through a familiar path.

3. It enlisted several engineers in the task of a large-scale tryout of the method over a wide range in machines, so that a far better correlation between test data and calculated data was obtained than would otherwise have been possible.

Some idea of the size of this program may be of interest here. The first program we used (described in reference 6 of the paper) contained 2,900 instructions, 290 program constants, and 360 items of data. After we had made these calculations on some 1,500 to 2,000 motor designs and had studied the results, we developed a newer and better procedure, which also included double-cage induction motors; we then rewrote the procedure entirely from scratch. The new procedure contained some 3,500 instructions, 350 program constants, and 270 items of data. These figures show why we would not have considered an IBM *650* computer adequate for this problem. It could have been done, of course, with the *650* by breaking the program up into parts and working with it in sections as is sometimes done for this type of problem. It is more convenient, however, to have the entire program on the drum.

We have done considerable work in different areas and our findings thoroughly support the conclusions of the authors that the use of synthesis is more desirable and more effective in a large-scale computer. In the synthesis procedure, we also recognize the need for flexibility. The authors refer to this as giving the engineer an opportunity to override the computer program. We would be interested in information as to how this is accomplished. One way is to duplicate some of the outputs on the input form. For example, in designing an induction motor, sometimes the engineer does not care whether the winding is connected in star or in delta; at other times he insists that it be connected one or the other way. We provide a space on the input blank for him to mark or leave blank. If he leaves the space blank, the machine operator types in a zero; the computer "looks at" the zero and then finds the better connection; if the engineer marks down one connection, then the computer will try only windings connected as instructed by the engineer.

To summarize briefly, our computer helps us to produce better and more economical designs of motors, faster. It gives us more output from the same number of engineers. Also, we find that, through the computer, beginners become productive sooner than without it.

REFERENCE

1. *Discussion* by **C. G. Veinott** of DIGITAL COMPUTERS AS AN AID IN ELECTRIC-MACHINE DESIGN, **R. M. Saunders.** *AIEE Transactions*, vol. 73, pt. I, May 1954, pp. 191–92.

Ned Chapin (Stanford Research Institute, Stanford, Calif.): As a statement of the authors' methods, I found the paper valuable and interesting. But I could not find a definite, concise statement of the "philosophy" in the title, unless the authors wish us to consider that a method constitutes a philosophy. It seems to me that it generally does not. May we have a paragraph clarifying the title of the paper?

Consider the following. Three design processes are listed: 1. "...cut-and-try." 2. "... to synthesize the design process," and 3. "automatic convergence." Ignoring the fact that these three are nonparallel in both their logic and their statement, the authors indicate that "design synthesis" (which apparently, although they do not so state, is the same as point 2 just cited) is what they have attempted on the automatic computer. However, the discussion of the method used appears to be what was previously referred to as "automatic convergence." And when the authors take up what they say they feel to be the major area of potential in the use of computers in design work (optimization) their illustration is taken from "cut-and-try." Would the authors clarify the philosophy mentioned in the title of the paper in relation to the implication in point four in the section "Evaluation"? Here, rather than the analytical approach, the authors appear to call for the use of a computer to tackle by "brute force" procedures too burdensome to accomplish by hand. If this is the intention, how do the authors relate it to their more general implication of identity of method and philosophy?

In the discussion of standardization, the authors do not specifically state what they wish to optimize. Yet the discussion of costs appears tacitly to assume that the measurable total cost to their company is to be optimized. Is this the intention?

I feel that the discussion of the effect of the degree of standardization on various types of costs is one of the most valuable parts of the paper, and I hope that the topic will be amplified in a future paper.

Harry N. Cantrell (General Electric Company, Schenectady, N. Y.): The authors are to be congratulated on their successful treatment of a difficult abstract subject. A careful study of the concepts discussed has been very rewarding.

This is a philosophy of the mechanization of engineering design work and as such is important to all interested engineers. It is not restricted to electric apparatus or even to electrical engineering but it does imply a design process that is repeated often enough to be worth mechanizing.

That it is possible and often desirable to mechanize engineering design work does not mean that engineers are obsolete. It does mean that the emphasis in engineering work is swinging toward more consideration of how to design a product and away from the actual design work itself. It also does not mean that engineering work is getting any easier. As the paper suggests, the problems involved in the application of digital computers to design are formidable engineering problems indeed. Furthermore, their solution requires the same thorough knowledge of the design process, the engineering fundamentals involved, and the relationships between parameters that characterize the experienced design engineer. This writer is convinced that only a good, experienced design engineer can define a mechanizable process for designing a product. This means that the engineer must learn a great deal about computers and more about mathematics and numerical analysis, but experience has shown that most good engineers have no great trouble in becoming proficient in these subjects.

Mechanizing engineering design also raises interesting management problems on such subjects as design responsibility, training of new engineers, and engineering organization, but these are beyond the scope of this discussion.

This writer has found the distinction between design analysis, design synthesis, and design optimization, as given in the paper, to be very helpful. These are really three very different stages of mechanization in which each stage is often several times more difficult to achieve than the preceding one.

D. J. Herder (General Electric Company, Lynn, Mass.): It is worthy of note that the authors have had the courage to tackle the problem of including manufacturing costs. While the inclusion of these costs is usually more difficult than the engineering problems involved, it is only by this means that the designer can utilize fully the capabilities of the high-speed digital computer for design optimization. The authors are to be especially complimented for this forward step. It would be interesting to have further details of their approach to the cost problem.

The authors break down design work into preliminary and final design. In either case the computer is used each time a design is required. In the case of preliminary design or advance design, the computer program may be used in another way which is sometimes overlooked: For advance design work, especially in determining the effect of new materials or new technological developments, quick, reasonably accurate hand methods are still of value to the designer. The decided advantage is that the logic is not buried in a maze of computer instructions. The computer can be an invaluable aid in this. Using design analysis programs, thousands of calculations may be made on the computer with many levels of the variables. The results may be properly plotted in such a way as to revolutionize the hand calculation methods for use in design synthesis procedures. While I agree that 2-dimensional graph paper has its limitations, it is still a valuable aid in visualizing the effects of variables. Only through such visualization can the designer's talents be fully developed and utilized.

As an example, in rotating machinery design, after the computer is used to grind out page after page of answers, it is possible to plot sets of curves of temperature rise versus ampere-turn loading for various rotor diameters, insulation thicknesses, and hydrogen pressures. With these curves the designer may choose his allowable temperature rise, enter the curves for the proper insulation thickness and hydrogen pressure, and read off the approximate rotor diameter required for any desired number of ampere-turns. Without the computer these curves would never exist because the time to prepare them by hand calculation would be prohibitive.

By repeated use of this approach, preliminary design synthesis can approach a series of reasonably straightforward steps which can be taken by the designer at his desk in a shorter period than is required to fill out input form sheets and to wait for the results from the computer, especially if the computer is used under a closed-shop arrangement.

P. A. Abetti and S. B. Williams: The authors wish to thank the discussers for their interest. We feel that the discussions, which are based on considerable knowledge and experience, add to the understanding of this relatively new and complex subject.

The question of which type of computer is best for design synthesis is a difficult one, and the answer depends largely on the degree of synthesis desired and the amount of detailed information to be generated. The authors spent several months studying this problem before the order was placed for an IBM *705*. While design synthesis is a more technical problem than payroll, the most difficult area of programming is not the calculation of performance, but rather the selection of the proper design data and the duplication of the decision-making process of the design engineer. Design synthesis of complex apparatus such as power transformers involves much rearrangement of data and maintenance of large tape files. The data fields in these files may vary from two or three characters to hundreds. The numbers used for calculations are rarely over four digits long. Confirming Dr. Grosch's observation, arithmetic instructions in many sections of our programs amount to considerably less than 30% of the total instructions. The manipulation of alphabetic information is a necessity if the machine is to communicate with humans. There is presently no computer commercially available that is ideally suited for comprehensive design synthesis work. Writing the specifications for such a machine would be difficult. In our opinion, the ideal computer would have very large random access memory, variable word length, externally fixed point and internally floating point arithmetic, a simple command structure, and would cost about one-tenth of its current price. After one year of operation, we still believe that the IBM *705* is the best available large computer for our particular problems. Rapidly changing computer technology may soon alter this picture.

Mr. Veinott's discussion of how he allows the design engineer to override the computer agrees exactly with the method which the authors and their colleagues have used. Before making a design choice the computer is told to "look at" a certain input position. If that position is blank, the computer chooses the proper data in what it considers to be the optimum fashion. If the input position is not blank, the machine uses the data found at that position. This feature is extremely useful for design studies of alternate constructions. Since it gives the engineer some control over the design, it also helps to allay the suspicion that the computer has taken over design responsibility.

The pitfalls that may appear when engineers try to discuss philosophies are very well pointed out by Mr. Chapin's discussion. We did not intend to imply that method and philosophy are the same. Rather, we meant to say that not only the methods by which we design apparatus are undergoing radical changes because of computers, but also that the broader aspects of the purpose of design and of the principles of design are changing. With a computerized design procedure available, we can afford to make many more designs and for the first time study the problems of optimization and standardization. The design engineer becomes an evaluator rather than a slide-rule pusher. Actually, the impact of computers is even more basic. Since the "brute force" approach of applying existing design systems to a computer is not economical, a complete change in the manner of codifying and tabulating design data may be necessary. The way in which design information flows through an engineering organization may be altered. Computers force simplification, systemization, and standardization of engineering design data and procedures.

Generally, the three design processes of analysis, synthesis and optimization are at different stages of sophistication, as Mr. Cantrell points out. To select one of these as most suitable is analogous to choosing among creeping, walking, and running. Without having gone through the first stage, it would be very difficult to tackle the second. Without having solved the second, the third would be impossible.

In the discussion of standardization, the authors intended to optimize the total cost of the product which benefits both manufacturer and user, while meeting all of the user's requirements and applicable standards.

In conclusion, we should like to summarize our philosophy of applying digital computers to the design of electric apparatus. The digital computer should not be used as a glorified slide-rule, to repeat without imagination what is being done by conventional methods. Rather it should be used according to more and more sophisticated methods, to do things which could not be done previously, such as complete design optimization and optimum standardization. Only in this manner will it be possible to take full advantage of this powerful new tool and to utilize completely the creative abilities of the design engineer.

Applications of Nonlinear Magnetics

HERBERT F. STORM
MEMBER AIEE

Synopsis: Quite deservedly, magnetic amplifiers have been given very wide publicity.[1-10] While much attention has been focused on the saturable reactor as a vital part, if not the only part, of a magnetic amplifier, the usefulness of the saturable reactor for other functions than amplification is not as widely appreciated.

It is the object of this paper to show that the very same saturable reactor, or, to put it more broadly, the very same saturable-core device, which is the heart of the magnetic amplifier, will produce important effects besides amplification. Such effects find utilization in counters, timers, voltage and current references, frequency multipliers, firing circuits for thyratrons and ignitrons, pulse-shaping circuits for magnetrons, rectifying circuits without conventional rectifiers, bistable flip-flops, frequency detectors, square-wave oscillators, and many others.

The method of presentation is first to consider the saturable-core device alone, and then in combination with other circuit elements such as resistors, linear inductors, or transistors. To facilitate absorption of the subject matter, only one component at a time will be introduced. The vast area of nonlinear magnetic devices, however, cannot be covered here completely. Therefore, the saturable-core device and other circuit components will be discussed only in combinations without repetitions; i.e., the circuits described contain only one representative of each type of circuit element. It will be seen that this method permits discussion of the fundamentals of most nonlinear magnetic devices with an irreducible minimum of components. Many references for further study will be found at the end of the paper.

IN ITS SIMPLEST FORM the saturable-core device consists of a ferromagnetic core with a single flux path, and one or several windings which are linked with essentially the same flux path (Fig. 1). This device will be designated by the abbreviation *SR* and no differentiation will be made between a saturable reactor, a saturating reactor, or a saturating transformer.

For didactic purposes, a toroidal core covered entirely by each winding is the preferred form. The ratio of outside to inside diameters of the core should not exceed 1.4. A suitable core material is anisotropic 50-50 nickel-iron, such as Orthonol or Deltamax.

The dots in Fig. 1 signify winding polarities. The voltage across each winding is defined as positive when the terminal with the dot is more positive than the other terminal. A current is called positive when it flows in the direction of the polarity arrow. The magnetomotive force is called positive when derived from a positive current.

Nomenclature

AT_c = coercive ampere-turns
E = supply voltage, volts, avg (average)*
E_G = gate voltage, volts, avg*
E_L = voltage on inductor L, volts, rms*
E_O = output voltage, volts, avg
E_m = maximum value of voltage, volts
E_C = capacitor voltage, volts, rms
E_R = voltage across resistor R, volts, avg*
$E_{G,C}$ = combined voltage across *SR* and C, volts, rms
e = supply voltage, volts, instantaneous
e_G = gate voltage, volts, instantaneous
e_O = output voltage, volts, instantaneous
e_L = voltage on inductor L, volts, instantaneous
e_R = voltage on resistor R, volts, instantaneous
f = supply frequency, cps (cycles per second)
I = load current, amps (amperes), rms
I_C = capacitor current, amps, rms
I_G = gate current, amps, rms
i = current, amps, instantaneous
N = number of turns
R_G = resistance of gate, ohms
R_L = resistance of load resistance, ohms
SR = saturable reactor, saturating reactor, saturable transformer
T = period of 1 cycle, sec (second)
t = time, sec
t_s = time of occurrence of saturation, sec
Φ_s = saturation flux, maxwells
$\Delta\Phi$ = flux change, maxwells
ϕ = flux, maxwells, instantaneous

SR Alone

Counter

Counting can be accomplished by utilizing the memory properties of a magnetic core. One form of a counting device, Fig. 2(A), consists of an *SR* which will energize the load after having received a predetermined number of pulses. The supply voltage e consists of pulses of constant volt-seconds, Fig. 2(B), whose number represents the number to be counted. The core properties are shown in Fig. 2(D). Before the count starts, the core has been reset to negative residual flux, as indicated by point O in Fig. 2(D). The first voltage (counting) pulse increases the core flux to a level indicated by 1 in Fig. 2(D), and the flux will remain at this level until the next voltage pulse arrives. As long as the core is unsaturated, the supply voltage appears almost exclusively across the terminals of the *SR*. This may be expressed as $e_G \approx e$. The flux $\Delta\Phi$ follows from

$$e_G = \frac{N}{10^8}\frac{d\phi}{dt} \tag{1}$$

$$\Delta\Phi \approx \frac{10^8}{N}\int e\,dt \tag{2}$$

During the presence of the voltage the load is traversed by a small magnetizing current, Fig. 2(C), determined by the flux-current loop of Fig. 2(D)

$$i = \frac{AT_c}{N} \tag{3}$$

Consecutive voltage pulses will raise the flux level by equal amounts. At the end of the ninth pulse, in this particular case, the core will be at positive residual flux, as indicated by point 9. When the next voltage pulse arrives, the core will be driven into saturation, point 10 in Fig. 2(D). Then the load current is no longer limited by equation 3, but increases substantially to

$$i = \frac{e}{R_L + R_G} \tag{4}$$

The large current pulse is the indication for a completed count; thus we have a counter of 10. The large current pulse can be used, for instance, to actuate a relay, or to provide an input pulse for a second stage of a decade counter.

The memory properties of magnetic cores are also utilized for the storage of information in large-scale computers.[11] For instance, in a binary system, each digit can have only two values, usually referred to as 0 and 1. With one residual flux assigned, e.g. the negative one to indicate 0, and the positive residual flux to indicate 1, the value of the digit can be stored. The time required to change the core from a 0 to a 1, or vice versa

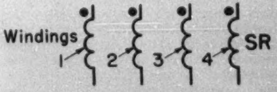


Fig. 1. Simple saturable-core device

Paper 58-180, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted November 4, 1957; made available for printing November 13, 1957.

Herbert F. Storm is with the General Electric Company, Schenectady, N. Y.

* In Figs. 8, 10, and 12, the voltages are measured in rms.

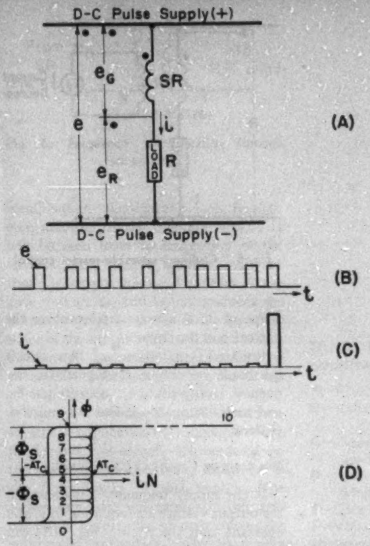


Fig. 2. Magnetic counter

A—Circuit diagram
B—Counting pulses
C—Current through load
D—Flux-current loop of counting core

(called flux-reversal time) limits the speed of the computer. For the purpose of minimizing the delay caused by eddy currents, the magnetically active part is often made of ferrite; ultrathin, vacuum deposited magnetic films promise a further reduction of flux reversal time.[12–14]

TIMER

If the counting pulses e, Fig. 2(B), arrive at a constant rate, the counter becomes a timer. Such a device has been used as a weld-control timer, counting up to 50 pulses per core.[15] A tap-changing switch, connecting various numbers of active turns, permits different time settings. Suitable core materials for counting applications are those with rectangular flux-current loops, such as Dynamax,[16] *65*-permalloy, 50-50 anisotropic nickel-iron (listed in sequence of preference). A pulse supply voltage of constant volt-seconds can be obtained from a circuit arrangement such as described in the following section, and shown in Figs. 3 and 4. A multistage, decade-counting device, which takes advantage of transistors, is described in reference 17.

ORDINARY SATURABLE-REACTOR CONTROL

When one winding of the *SR* is connected in series with the supply voltage and the load, and the other winding is connected to a direct-current source, the simplest form of a magnetic amplifier results (Fig. 5). This circuit, particularly in its double-core form, has found widespread attention in practice and literature,[1–8,18,19] but it falls outside the scope of this paper, as indicated in the synopsis. It will merely be mentioned that this circuit is not only used as a power amplifier, but also as a constant-current source, a-c limiting reactor, phase shifter, frequency doubler, and measuring device for large direct currents and high direct voltages. While the above list of properties is also exhibited to some extent by the single-core arrangement of Fig. 5, the greater efficacy of the double-core arrangement makes it the preferred form for practical applications.

CORES WITH MULTIPLE-FLUX PATHS

Since cores with multiple-flux paths (e.g. Transfluxor) are outside the scope of this paper their discussion will be found in references 20 through 22.

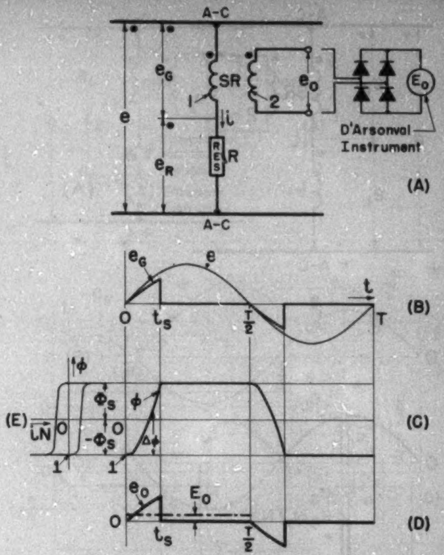


Fig. 3. Voltage or current reference circuit

A—Circuit diagram
B—Supply voltage e and gate voltage e_G
C—Core flux
D—Output voltage e_O
E—Flux-current loop

SR+Resistor R

FREQUENCY DETECTOR

Let us connect an *SR* across an a-c bus, Fig. 3(A). The resistance of winding 1 is lumped into *R*. Suppose the core flux ϕ is at negative saturation at $t=0$ (points 1, Figs. 3(C) and (E). For $t>0$, the supply voltage e becomes positive, Fig. 3(B), and the core leaves negative saturation. The core flux increases according to equation 1. A magnetizing current will flow, causing a voltage drop on the circuit resistance *R*, thus the gate voltage e_G may be smaller than the supply voltage e. Hence,

$$\Delta\Phi=\frac{10^8}{N}\int e_G dt \tag{5}$$

The output voltage e_O, Fig. 3(D), follows from

$$e_O=e_G N_2/N_1 \tag{6}$$

The average output voltage E_O for the first half-cycle is

$$E_O=\frac{2}{T}\int_0^{\frac{T}{2}} e_O dt \tag{7}$$

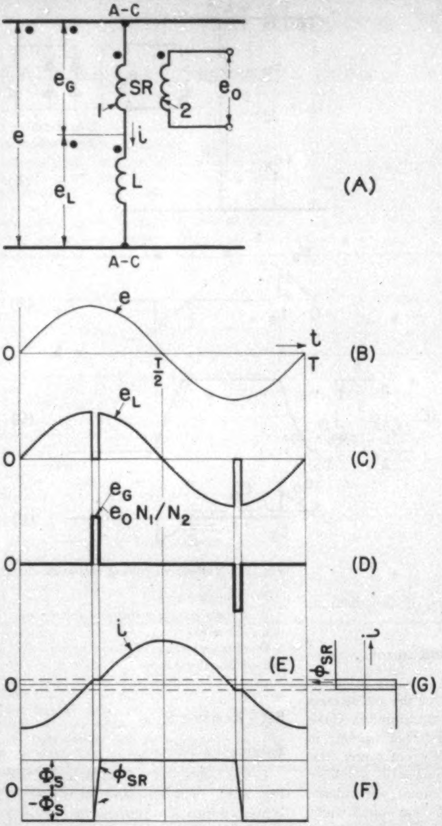


Fig. 4. Peaking circuit

A—Circuit diagram
B—Supply voltage
C—Voltage on inductor L
D—Gate and output voltages
E—Current in winding 1 of SR
F—Flux in SR
G—Flux - current loop for core of SR

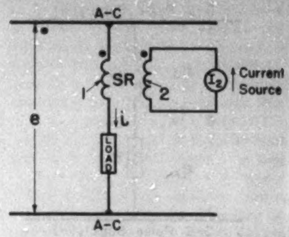


Fig. 5. Ordinary saturable-reactor control

At $t=t_s$, the core is said to saturate, $\Delta\Phi=2\Phi_s$, Fig. 3(C), and it will be assumed that the core will remain saturated for the remainder of the half-cycle. Hence

$$E_O=\frac{2}{T}\;\frac{N_2}{N_1}\int_0^{t_s} e_G dt \tag{8}$$

Substituting from equation 5 into equation 8, and for $\Delta\Phi=2\Phi_s$,

$$E_O=4f\Phi_s N_2 10^{-8} \tag{9}$$

where the supply frequency f replaces the period T. The events during the second half-cycle produce similar results. A D'Arsonval instrument energized through a loss-free rectifier, Fig. 3(A), will read E_O.

One concludes that the output voltage E_O is proportional to the supply frequency. The output voltage E_O is not affected by the magnitude of the supply voltage, as long as the core saturates at least once in each half-cycle. The wave shape of the supply voltage e is irrelevant as long as it does not go through zero more than once per half-cycle. The magnitude of the resistance R does not affect E_O, it merely controls the maximum value of current i (equation 4). If a current flows in the output circuit, the output voltage E_O is only reduced by the voltage drop in winding *2* of the *SR*, but is not affected by the voltage drop in *R*. As long as the saturation branches of the flux-current loop, Fig. 3(E) are horizontal in the first and third quadrants, the width of the flux-current loop and the shape of its flanks are irrelevant to the operation of the circuit.

Practical applications of this circuit are made for frequency regulators, frequency compensators, over-frequency, and under-frequency indicators and controllers.

VOLTAGE OR CURRENT REFERENCE

If the supply frequency f is constant, the output voltage E_O (equation 9) is also constant, and the circuit, Fig. 3(A) can be used as the basis for a voltage or current reference circuit.[23] If the supply frequency is variable, frequency compensation can be applied, by, for instance the addition of linear reactors in the output circuit.[24,25] Reference 25 describes a circuit which holds the output current at 15±.02 ma (milliamperes) d-c, while the load is 50 ohms, the supply voltage varies from 100 to 140 volts, the supply frequency varies from 360 to 440 cps, and the temperature varies from −55 to 100 C (degrees centigrade). The device has a volume of 15 cubic inches and weighs 1.5 pounds.

FIRING CIRCUITS FOR THYRATRONS

For the precise firing of grid-controlled mercury arc or gaseous rectifiers, a rapid rise of the grid-cathode voltage is needed. Such a voltage develops across the resistor *R*, Fig. 3, at the instant t_s of core saturation. This voltage e_R can be applied directly, or by interposition of a so-called grid transformer to the grid-cathode circuitry.[26]

SR+Linear Inductor L

PEAKING CIRCUIT, FREQUENCY MULTIPLIER

Let us replace the resistor *R* of Fig. 3(A) by a linear inductor *L*, as shown in Fig. 4(A). In order to emphasize the fundamental action of this circuit, let us assume zero resistance for all circuits, and an

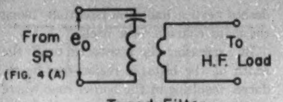


Fig. 6. Frequency multiplication through tuned filter

ideal, square flux-current loop for the core material of the *SR*, Fig. 4(G). It will be seen from an inspection of the various wave shapes shown in Fig. 4 that the supply voltage e appears as the gate voltage e_G, and hence produces the output voltage e_O at or near the crest value of the supply voltage e, thus making this circuit far more effective for the generation of short pulses than the circuit of Fig. 3(A).

The output voltage E_O may be used for a variety of applications, for example where fixed input volt-seconds are needed (viz. the counter and the timer, described in the first section of this paper). The pulses can also be used for the grid control of thyratrons. Furthermore, the substantial content of higher harmonics of the pulses can be utilized for frequency multiplication. By interposing a tuned filter between the *SR* and the load, as shown in Fig. 6, a current having the resonance frequency of the filter will flow in the load circuit. It is obvious from the function of the circuit that the filter is shock-excited, and hence the high frequency voltage will be of damped character. This method of frequency multiplication was used in the early days of radio telegraphy, where supply frequencies of 5 to 10 kc were available, which were stepped up by a factor of 15.

Connecting a linear capacitor in series with the linear inductor L of Fig. 4(A) increases the circuit efficiency to 50% when multiplying the supply frequency by a factor of 47.[27]

Other methods of frequency multiplication involve multiple magnetic cores, and hence fall outside the scope of this paper. For the reader who is interested in frequency multipliers, see references 28 through 34.

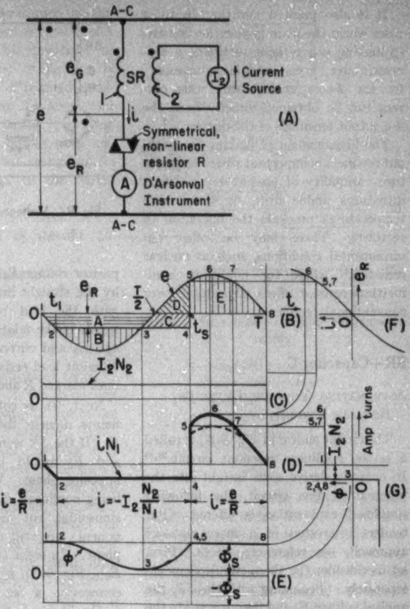


Fig. 7. Direct current without rectifiers

A—Circuit diagram
B—Voltage inventory
C—Ampere-turns in winding 2
D—Ampere-turns in winding 1
E—Core flux
F—Voltage current characteristic of non-linear, symmetrical resistor R
G—Flux-current relation

SR+Nonlinear, Symmetrical Resistor

DIRECT CURRENT WITHOUT RECTIFIERS

The circuit consists of an *SR*, Fig. 7(A), whose control circuit is energized from a d-c source I_2. A symmetrical, nonlinear resistor is connected in series with the gate circuit.[35] A suitable nonlinear, symmetrical resistor may be made of carbon, or silicon-carbide such as is known as Thyrite.[36] The provision of a current source, rather than a voltage source, for the energization of winding 2 leads to a smooth current I_2 as shown in Fig. 7(C). The action of the load current i can be traced from the Figs. 7(D) and (G). (A detailed description of a related circuit action can be found in Chapter 8 of reference 8). The core saturates at t_s; that is, when the flux upswing equals the flux down-swing, Fig. 7(E). One sees from Fig. 7(B) that areas

$$B=C+D \tag{10}$$

During the interval $t_s<t<T$, the core is saturated; hence $e_R=e$. The volt-seconds acting on the nonlinear, symmetrical resistor are expressed by area E

$$E=A+B-D \tag{11}$$

Substituting from equation 10

$$E=A+C \tag{12}$$

One concludes that the positive volt-seconds (E) from $t_s<t<T$ equal the negative volt-seconds ($A+C$) acting on the load from $0<t<t_s$. Then, for a linear load resistance, the resultant positive-load ampere-seconds (dotted section to 8 of Fig. 7(D)) must equal the negative-load ampere-seconds (1, 2, 4), and hence the load current is a pure alternating. If, however, the load resistor is nonlinear, although symmetrical, as shown by its characteristics in Fig. 7(F), the load resistance will in general be smaller during the interval from t_s to T than from O to t_s. As a result, the positive-load ampere-seconds are now larger than the negative-load ampere-seconds, and hence the load current contains a d-c component. The D'Arsonval ammeter shown in Fig. 7(A) will indicate current flow. If the direction of current I_2 is reversed, the d-c component of the load current will also be reversed.

A gradual change of I_2 causes a gradual change of the d-c component of the load current, and hence, the circuit can also be used for amplification purposes.

In addition to the d-c component, the load current also contains a-c components. The ratio of d-c components to a-c components can be greatly increased by using two *SR*s and two nonlinear, symmetrical resistors feeding into a common load. Among other usages, this arrangement has been applied to a magnetometer; see page 466 of reference 8.

It is also pointed out that in those cases where the main application consists in making d-c without rectifiers, a separate source for current I_2 is not necessary for the 2-core arrangement: this current can be obtained directly from the d-c output terminals of the device.

The phenomenon of making d-c without rectifiers is important where temperature variability of present-day rectifiers introduces undue drift, or where high temperatures preclude the operation of rectifiers. There may be other environmental conditions, such as nuclear radiation[37] where the nonlinear, symmetrical resistor offers advantages over conventional rectifiers.

SR+Capacitor C

ALTERNATING VOLTAGE DETECTOR, BISTABLE CIRCUIT

The circuit shown in Fig. 8(A) is called a series, nonlinear resonant circuit.[38,39] In order to derive some insight into the function of this circuit, the following simplified explanation is offered. (For readers interested in a more rigorous approach, see references 40–43.) First let us consider the three circuit elements separately. Beginning with the *SR*, the applied voltage E_G and the current I to the *SR* are related by the curve marked *SR*, Fig. 8(B); current I and all voltages in Fig. 8(B) are expressed in effective (rms) values. Next, the relation between capacitor voltage E_C and current I is shown by the straight line marked *C*. Finally, with the load being assumed purely resistive, the relation between load voltage E_R and current I is shown for two different load resistances by the straight lines marked *R* and *R′*.

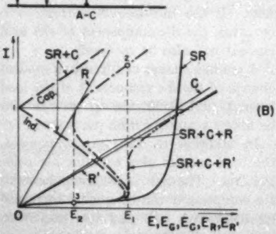


Fig. 8. Ferroresonance

A—Series, nonlinear resonant circuit
B—Voltage and current relations

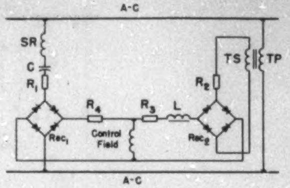


Fig. 9. Voltage-error detecting circuit

Next, let us combine two circuit elements, namely, the *SR* and the capacitor *C*. If the *SR* were a linear inductor, E_G and E_C would be sinusoidal and of opposite phase. With the *SR* actually being nonlinear, E_G and E_C are no longer sinusoidal, but their fundamental components are still of essentially opposite phase. In order to establish the voltage $E_{G,C}$, Fig. 8(A), which is across the series connection of the *SR* and the capacitor *C*, the liberty is taken of obtaining $E_{G,C}$ by subtracting E_C from E_G, resulting in the dashed characteristic labeled *SR+C* in Fig. 8(B). If the current is smaller than is indicated by the distance 0-1, it is inductive; if larger than 0-1, it is capacitive with respect to $E_{G,C}$.

The terminal voltage E is approximated by the vector sum of $E_{G,C}$ and E_R, resulting in $E=\sqrt{E_{G,C}^2+E_R^2}$. Thus, the characteristic marked *SR+C+R* has been determined in Fig. 8(B).

It will be noticed that the characteristic *SR+C+R* is bistable; i.e., when the supply voltage E is increased from zero, the current increases according to the lower branch of the characteristic. If the supply voltage is increased merely a trifle beyond E_1, the load current will increase to point 2. For a further increase of supply voltage E, the load current follows the upper branch of the characteristic. If the supply voltage is now decreased down to E_2, the current will decline according to the upper branch of the control characteristic. However, a minute decrease of the supply voltage E below E_2 will cause the load current to establish itself at point 3. For supply voltages between E_1 and E_2, two stable states for the load current are possible, hence the name bistable circuit. The bistable region is not traversed along the dashed vertical lines, but essentially along the main characteristic itself.

By increasing the resistor R to R' the width of the bistable region can be reduced, resulting in the border case where the loop width becomes zero and the characteristic becomes vertical (the curve marked *SR+C+R′*). With a further increase of load resistance, the characteristic remains single-valued, and its slope decreases.

The steep part of the characteristic (*SR+C+R′*) makes the circuit a useful tool for sensing deviations of the alternating supply voltage. For instance, such sensing is utilized in voltage regulators for alternators (see reference 44 and reference 8, p. 480).

One form of application to an alternator voltage regulator[45] is shown in Fig. 9. The control field (to be acted upon by the nonlinear, resonant circuit) may belong to a rotating exciter or a magnetic amplifier which directly or indirectly energizes the field of the alternator. The current flowing through the control field is the difference between a current proportional to the line voltage, and another current coming from the nonlinear resonant circuit, which sharply varies as the supply voltage varies.

The bistable characteristic, *SR+C+R*, Fig. 8(B), is utilized for the handling of logic in digital systems. When this is combined with a second circuit, identical to that of Fig. 8, a trigger pair is obtained, which forms the basic building block for binary counters, storage registers

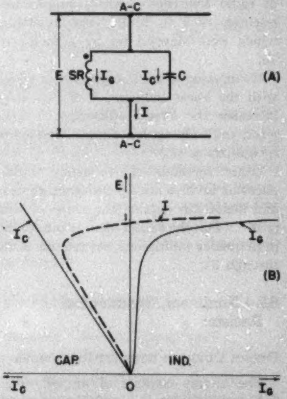


Fig. 10. Ferroresonance

A—Parallel, nonlinear resonant circuit
B—Voltage, current, and phase angle relations

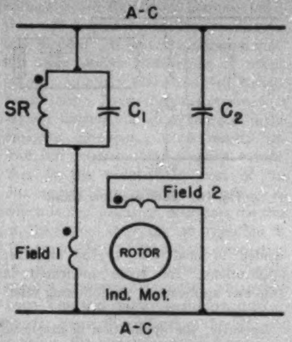


Fig. 11. Reversible induction motor drive

and other logic operations.[46] Triggering is usually achieved by single-polarity pulses which produce simultaneously mmfs (magnetomotive forces) in both cores.

REVERSIBLE INDUCTION MOTOR DRIVE

The circuit of Fig. 10(A) is called a parallel, nonlinear, resonant circuit. The currents through the *SR* and the capacitor *C* are shown in Fig. 10(B). Because of the phase opposition between I_θ and I_C, the total current I is assumed to be their difference. If the supply voltage E is larger than is indicated by the distance 1-0, the current I will be inductive; if E is smaller than 1-0, the current I will be capacitive.

Reversible, induction motor drives are often used as control means for line voltage regulators, where they may actuate a tap-changing mechanism of a transformer, or they may actuate an induction-type voltage regulator. If one field of the control motor is connected in series with a parallel, nonlinear resonant circuit (SR-C_1, Fig. 11), and the other motor field is connected in series with a power-factor correction-capacitor C_2, the motor will rotate whenever the supply voltage deviates from the set point. With proper phasing, the motor armature can be made to rotate in such a direction as to have the tap-changing transformer or induction regulator restore the supply voltage to the set point.

FLASHING LIGHTS, SUBHARMONIC OSCILLATOR

In reference, again, to the circuit of Fig. 8(A), if a d-c bias is applied to another winding of the *SR* (not shown), the circuit may break into subharmonic oscillations.[47] The period of oscillation can be varied by variation of the bias over a range of from several cycles to several seconds. This tendency to break into subharmonic oscillations is useful in flashing lights of railroad-crossing signals, or flashing beacons, and other applications requiring pulsating energization. As this circuit uses exclusively static means for switching, its maintenance is reduced to the barest minimum.

For further study of subharmonic oscillators see references 48 through 50.

SR+Capacitor C+Linear Reactor L

LINE VOLTAGE STABILIZER, FREQUENCY SENSOR

As long as the *SR* saturates twice in each cycle (see section on the frequency detector), the average gate voltage across the *SR* will remain fairly constant (equation 9). When the load resistance R is connected in parallel to the *SR* as shown in Fig. 12, the average value of the load voltage E_R will also remain fairly constant. In order that the difference between the variable supply voltage E and the average load voltage E_R be taken up, a linear inductor L is provided for.[51]

As indicated, the load voltage E_R is not perfectly constant, but it will change slightly in the same direction as the supply voltage changes. In order to counteract this variation, a voltage is added to the load circuit, which will decrease the load voltage when the supply voltage increases, and vice versa. Such a voltage can be introduced into the load circuit by connecting the load to tap *2* of the linear inductor L, instead of to tap *1*. The capacitor C improves the regulation properties of the circuit, and affects the wave shape of the output voltage.

Line voltage stabilizers, using the above principle, are commercially available for output powers of a few watts or up to many kilowatts. The output voltage usually is held within ±1% of rated voltage for supplying voltage variations of ±15%. In most commercial stabilizers the output circuit is isolated from the supply circuit. The speed of response of the circuit is usually such as to correct for line voltage fluctuation in less than two cycles. Another important feature of this circuit is the inherent limitation of line currents when there are short circuits in the load circuit.

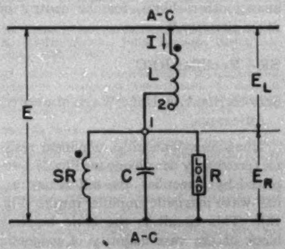


Fig. 12. Basic circuit for line voltage stabilizer

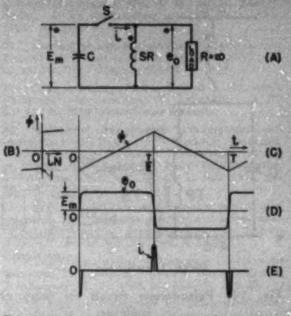


Fig. 13. Square-wave constant-voltage power supply

A—Simplified circuit part
B—Flux-current relation
C—Core flux
D—Output voltage
E—Current between C and SR

As the supply frequency changes, the number of pulses of fixed volt-seconds, as controlled by the *SR*, will also change; hence the load voltage becomes frequency-sensitive (equation 9). In cases where this sensitivity is undesirable, frequency compensators can be supplied.

However, there are applications, as discussed under "Frequency Detector," where the object is to indicate or control the supply frequency. In these cases the load current of the voltage stabilizer is compared with a current derived directly from the same voltage via a fixed impedance, the difference between these currents being the error signal. A circuit similar to that shown in Fig. 9 would be useful. For a fixed supply voltage, the output voltage of the stabilizer will increase by about 1.5% for an increase of 1% of the rated supply frequency.

SQUARE-WAVE, CONSTANT-VOLTAGE POWER SUPPLY

The designer of a voltage stabilizer has several means of affecting the wave shape of the output voltage. In most

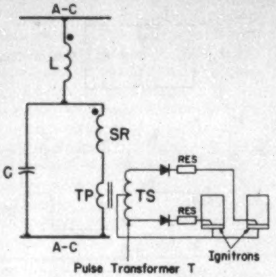


Fig. 14. Pulse-former circuit for firing of ignitrons

cases he will attempt to obtain a sine shape. There are, however, cases where a square voltage is more desirable, one in point being a power supply for magnetic amplifiers whose control characteristics should be inherently linear.[52]

The output voltage E_R (Fig. 12) can be made fairly square by using a square-loop core material and toroidal construction for the *SR*. The following is intended as a simplified explanation of the function of the circuit.[53,54] The core material of the *SR*, Fig. 13(A), is represented by the flux-current relation shown in Fig. 13(B). The load resistance *R* is assumed to be infinitely large, the switch *S* is open, and capacitor *C* has been charged to a potential of E_m volts. The core of the *SR* has been set to point *1* of Fig. 13(B).

Let us now close the switch *S* at $t=0$. The voltage E_m is now applied to the *SR*, and its core flux ϕ increases according to

$$\phi=-\Phi_s+\frac{10^8}{N}\int E_m dt \qquad (13)$$

By assuming zero width of the flux-current loop of the core, and infinite resistance for the load, no current will be drained from the capacitor *C;* hence the output voltage $e_0=E_m$. The flux ϕ increases linearly, as shown in Fig. 13(C). At $t=T/2$, saturation is reached. The only impedance in the discharge path of the capacitor *C* is assumed to be derived from the air flux of the *SR*. A sinusoidal current *i* of very short duration will flow, Fig. 13(E), transferring the electrostatic energy of the capacitor into electromagnetic energy of the air field of the reactor, and then returning this air field energy to the capacitor *C*, which results in a voltage reversal of the capacitor. Under the influence of $-E_m$ the core of the *SR* will now decrease its flux linearly until $t=T$, where another reversal, similar to the one described above, takes place. Since the losses for this circuit are assumed to be zero, this oscillation will continue, resulting in an output voltage of e_0 of nearly square-wave shape.

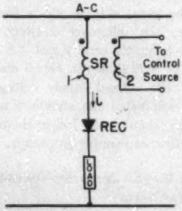


Fig. 15. Self-saturated half-wave magnetic amplifier

If the circuit of Fig. 13 is connected, by means of a linear inductor *L*, to an alternating supply voltage (resulting in the circuit previously shown in Fig. 12), continuous generation of square waves is assured. At the same time, the constant voltage properties of this circuit are retained. It should be pointed out, however, that the flow of output current into the load, causes a deterioration of the squareness of the output voltage.[53,54]

Pulse-Former Circuit, Firing of Ignitrons

It has been shown in Fig. 13(E) that a sharp current pulse flows between capacitor *C* and *SR*. This pulse can be utilized for the firing of ignitrons, as shown in Fig. 14. The circuit will be recognized as derived from Fig. 12, by a connection of the primary *TP* of a pulse transformer in series with *SR*. The secondary *TS* of the pulse transformer energizes the ignitors via rectifiers and resistors.[55] Rectifiers are used to avoid damage of the ignitors by reverse current.

The circuit shown in Fig. 12 also serves as the starting point for so-called "pulse-sharpening circuits," which are used, among other things, for the control of radar circuitry.[55]

SR+Rectifier REC

Self-Saturating Half-Wave Magnetic Amplifier

When the symmetrical, nonlinear resistor (previously described; see Fig. 7) is replaced by a rectifier, the self-saturating, half-wave magnetic amplifier results (Fig. 15). This circuit is the basic building block of the vast majority of magnetic amplifier applications.[56-60]

As stated in the synopsis, it is not intended to deal with typical amplifier applications; for those interested in amplifier applications the following references are listed: 1–10, 61, and 62.

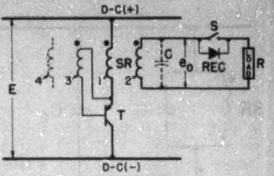


Fig. 16. Basic oscillator circuit

Recently, the application of magnetic amplifiers to static switching applications has been revived. Since the operation of many of these switching applications is analogous to typical magnetic-amplifier operation, merely references to static switching applications are listed here: 39, 47, 63–68.

Cores and rectifiers are also used increasingly in computers and in digital information-handling equipment. As these circuits usually use a plurality of cores and rectifiers their discussion falls outside the scope of this paper. For the reader who is interested in this field, references 69 through 73 have been prepared.

SR+Transistor T

Oscillator, Frequency Modulator

An *SR* is connected in series with a transistor *T* as shown in Fig. 16. The winding *2* of the *SR* is connected to the load resistor by means of the closed switch *S*.[74]

Let us assume that the core of the *SR* is unsaturated. If the direct supply voltage is now applied to the circuit, the transistor will pass some current, and part of the direct supply voltage will appear as a positive voltage across winding *1* of the *SR*. This voltage is transformed into winding *3*, which provides positive feedback and makes the transistor fully conducting. A positive output voltage e_0 is induced simultaneously in winding *2*, and also charges the distributed winding capacitance of the *SR*, indicated by *C*.

When the core flux reaches the positive knee of the flux-current loop, the voltage induced into winding *3* will become smaller, causing the emitter-collector resistance of the transistor to go up. This process is cumulative to the point where the voltage across winding *1* will even reverse (that is, become negative), in order to maintain its state of magnetization according to Lenz' Law. The

Vol.
Pt. 1
1958

voltage on winding *3* also becomes negative, and current flow in the transistor is now blocked. However, a negative voltage also appears on winding *2*, which is able to carry a current to satisfy Lenz' Law. Thus, the capacitance C is rapidly discharged, and recharged in the opposite direction. As a result, the core flux returns from positive saturation to the unsaturated part of the flux-current loop. The output voltage e_0, which is now negative, decreases the flux in the cores according to equation 2. The transistor is blocked from conduction by winding *3*. The core flux finally reaches negative saturation, whereupon Lenz' Law will cause a positive voltage to appear across all windings of the *SR*. As a result, the transistor again turns fully on, and completes the cycle of operation, producing alternating current in the load resistor R.

If a series of unidirectional pulses in the load resistor is desired, switch S is opened, and thus rectifier *REC* becomes effective.

The frequency of the oscillator can be controlled over a range of 3 to 1 by varying the direct supply voltage E.[74] A different manner of varying the frequency consists in introducing a d-c control signal to the fourth winding of the *SR*.

An *SR* can be combined with two transistors to form a square-wave oscillator, which, among many applications, can be used to energize magnetic amplifiers.[75–79]

Combinations of transistors and magnetic cores are used in computer engineering.[71,80,81]

Conclusions

The combinations of a saturable reactor with other circuit components produce extraordinary effects which, in general, cannot be duplicated by linear circuitry. While many of these effects have been utilized in the past, it is the hope of the author that this paper will stimulate further study, development, and utilization of nonlinear magnetic circuitry.

References

1. Saturating Core Devices (book), L. R. Crow. Scientific Book Publishing Company, Vincennes, Ind., 1949.

2. Magnetic Amplifiers (book), G. M. Ettinger. John Wiley & Sons, Inc., New York, N. Y., 1953.

3. Magnetic Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

4. The Transductor Amplifier (book), U. H. Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947.

5. The Transductor (book), A. U. Lamm. Esselte Aktiebolaget, Stockholm, Sweden, second edition, 1948.

6. The Magnetic Amplifier (book), J. H. Reyner. Stuart and Richards, London, England, 1950.

7. Magnetic Amplifiers and Saturable Reactors (book), M. G. Say. Newnes, London, England, 1954.

8. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., 1955.

9. Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art, James G. Miles. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2104–23.

10. Types of Magnetic Amplifiers—Survey, James G. Miles. *Ibid.*, vol. 71, pt. I, July, 1952, pp. 229–38.

11. Magnetic and Ferro-electric Computing Components, V. L. Newhouse. *Electronic Engineering*, London, England, May 1954, pp. 192–99.

12. Ferrite Apertured Plate for Random-Access Memory, J. A. Rajchman. Proceedings of the Eastern Joint AIEE, ACM, IRE Computer Conference, *AIEE Technical Publication T92*, May 1957, pp. 107–15.

13. A Compact Coincident-Current Memory, A. V. Pohm, S. M. Rubens. *Ibid.*, pp. 120–23.

14. On the Minimum of Magnetization Reversal Time, R. Kikuchi. *Journal of Applied Physics*, New York, N. Y., vol. 27, no. 11, Nov. 1956, pp. 1352–57.

15. Die Zähldrossel, P. Duffling, M. Tschermak. *Siemens-Zeitschrift*, Erlangen, Germany, Apr. 1952, pp. 140–44.

16. Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material, G. H. Howe. *AIEE Transactions*, vol. 75, pt. I, Nov. 1956, pp. 548–51.

17. A High-Accuracy Static Time Delay Device Utilizing Transistors, G. F. Pittman, Jr. *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 54–58.

18. The Transductor and Its Applications, A. U. Lamm. *ASEA Journal*, Altmänna Svenska Aktiebolaget, Vasteras, Sweden, vol. 31, 1939, pp. 66–80.

19. The Design of Transductor Circuits. *Ibid.*, 1948, pp. 3–20.

20. The Transfluxor, J. A. Rajchman, A. W. Lo. *Proceedings*, Institute of Radio Engineers, New York, N. Y., 1956, pp. 321–32.

21. Theory of Magnetic Cross Valves, W. H. Higa. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 204–08.

22. Transductors with Four-Limbed Cores, A. G. Milnes. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. 2, no. 83, Oct. 1954, pp. 554–58.

23. Selbsttätige Strom- und Spannungs-Regler, W. Geyger. *Archiv für Technisches Messen*, Munich, Germany, *JO62-7*, Nov. 1934; *JO62-8*, Dec. 1934.

24. A Saturable-Core-Current Reference Circuit for Use with Magnetic Amplifiers, A. G. Milnes, T. V. Vernon. *Technical Note no. EL. 46*, Royal Aircraft Establishment, Farnborough, England, Oct. 1952.

25. Saturating Transformer Reference Circuit, W. G. Evans. *AIEE Transactions*, vol. 73, pt. I, 1954, (Jan. 1955 section) pp. 610–16.

26. Theory der Stromrichter (book), A. Glaser, K. Müller-Lübeck. Julius Springer, Berlin, Germany, 1935, p. 237.

27. Ein neuer Hochfrequenz Maschinensender für Drahtlose Telegraphie, K. Schmidt. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 40, Oct. 4, 1923, pp. 910–14.

28. An Introduction to the Theory of Magnetic Frequency Multipliers Using Biased Magnetic Cores, D. W. Leiby. *M.S. Thesis*, Union College, Schenectady, N. Y., 1956.

29. Magnetic Frequency Conversion, L. C. Harriott. *Proceedings*, National Electronics Conference, Chicago, Ill., 1953.

30. Magnetic Frequency Multiplier for Fluorescent Lighting, E. G. Downie. *Electronics*, New York, N. Y., Mny 1955, p. 224.

31. Analysis and Design of a Magnetic Frequency Multiplier, Otto J. M. Smith, Jalal T. Salihi. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 99–106.

32. Design of the Resistively Loaded Static Frequency Doubler, P. P. Biringer. *Ibid.*, vol. 75, pt. III, Oct. 1956, pp. 834–39.

33. Characteristic Properties of a Magnetic Frequency Multiplier, W. Leonhard. *AIEE CP 56-716*, 1956 (available on request).

34. Magnetic Frequency Multipliers and their Rating, Part I—Frequency Triplers, William McMurray. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 384–90; Part II—Frequency Sextuplers, *Ibid.*, vol. 76, pt. I, July 1957, pp. 289–93.

35. Einrichtung zum Umwandeln von Wechselstrom in Gleichstrommittels gleichstromvormagnetisierter Drosselpulen, W. Schott. *German Patent no. 546,269*, Dec. 3, 1925.

36. The Calculation of Circuits Containing Thyrite, T. Brownlee. *General Electric Review*, Schenectady, N. Y., 1934, pp. 175–79; 218–23.

37. Effect of Radiation on Semiconductors' L. C. Pigg, C. C. Robinson. *Electrical Manufacturing*, New York, N. Y., Apr. 1957, pp. 116–24.

38. Studies in Non-Linear Circuits, Chauncey Guy Suits. *AIEE Transactions*, vol. 50, June 1931, pp. 724–36.

39. New Applications of Non-Linear Circuits to Relay and Control Problems, C. G. Suits. *Ibid.*, vol. 51, Dec. 1932, pp. 914–22.

40. Non-linear Mechanics (book), N. Minorski. J. W. Edwards, Ann Arbor, Mich., Mar. 1947.

41. Oscillatory Circuits Containing Iron-Cored Inductances, Abd El-Samie Mostafa, M. El-Shishini. *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 274–83.

42. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Jalal T. Salihi. *Ibid.*, vol. 75, pt. I, July 1956, pp. 296–307.

43. Series Circuit Incorporating a Non-linear Reactor, R. P. Evans. *AIEE CP no. 56-683*, (available on request).

44. An Alternator-Voltage Regulator Utilizing a Nonlinear Circuit, Harry W. Mayne. *Ibid.*, vol. 56, Apr. 1937, pp. 462–64.

45. Development of a Modern Amplidyne Voltage Regulator for Large Turbine Generators, W. A. Hunter, M. Temoshok. *Ibid.*, vol. 71, pt. III, Oct. 1952, pp. 894–901.

46. The Ferroresonant Trigger Pair: Analysis and Design, Charles F. Spitzer. *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 407–16.

47. Flashing Lamps Without Moving Parts' C. G. Suits. *Electrical World*, New York, N. Y June 25, 1932.

48. Recherches sur la démultiplication de fréquence ferromagnétique, R. Dehors. *Revue Générale de L'Électricité*, Paris, France, Nov. 1947, pp. 455–67.

49. Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis, Egon Brenner. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 450–56.

50. Circuit for a Magnetic Subharmonic Pulser, K. C. Hu, Y. H. Ku. *Ibid.*, vol. 76, pt. I, May 1957, pp. 137–41.

51. Anordnung zur Entnahme Konstanter Verbraucher-Spannung aus einem Netz veränderlicher Spannung, M. Osnos. *German Patent no. 578,321*, Oct. 21, 1925.

52. A Magnetic Trip Amplifier, R. B. Hodson, E. W. Manteuffel. *AIEE CP 56-292*, 1956 (available from the authors).

53. A Magnetic Voltage Reference, E. Manteuffel, R. O. McCrary. *AIEE Technical Publication T-86* "Conference on Magnetic Amplifiers." Joint AIEE, IRE, ISA Report, July 1956, pp. 200–10.

54. Oscillating Circuit Incorporating a Choke with Rectangular Magnetization Curve, T. Buchhold, E. Stuhlinger. *Technical Report no. 70*, Redstone Arsenal, Huntsville, Ala., Jan. 1952.

55. The Use of Saturable Reactors as Discharge Devices for Pulse Generators, W. S.

Melville. *Proceedings*, Institute of Radio Engineers, vol. 98, pt. III, no. 53, May 1951, pp. 185–207.

56. The Single-Core Magnetic Amplifier as a Computer Element, **R. A. Ramey.** *AIEE Transactions*, vol. 71, pt. I, 1952, (Jan. 1953 section) pp. 442–46.

57. The Biased Rectifier Amplifier—A Pulse Magnetic Amplifier, **R. E. Morgan.** *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 584–90.

58. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, **Harold W. Lord.** *Ibid.*, vol. 72, pt. I, 1953, (Jan. 1954 section), pp. 721–28.

59. Magnetic Amplifiers Serve to Teach Power Modulators, **H. C. Bourne, R. M. Saunders.** *Electrical Engineering*, vol. 75, Oct. 1956, pp. 943–47.

60. On the Input Power of the Half-Wave Magnetic Amplifier Circuit, **T. Kikuchi, K. Murakami.** *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 10–14.

61. Magnetic Amplifiers in Industrial Control Systems, **S. J. Campbell.** *Electrical Manufacturing*, New York, N. Y., Nov., Dec. 1955.

62. Applying Magnetic Amplifiers, **W. J. Dornhoefer, V. H. Krummenacher.** *Ibid.*, vol. 47, no. 3, Mar. 1951, p. 94; Apr. 1951, p. 112; vol. 48, no. 2, Aug. 1951, p. 106; Sept. 1951, p. 92.

63. Cypak Systems and Applications of Logic Functions, **J. P. Baker.** *Westinghouse Engineer*, Pittsburgh, Pa., July 1956, pp. 107–13.

64. Magnetic Logic Circuits for Industrial Control Sytstms, **W. G. Evans, W. G. Hall, R. I. Van Nice.** *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 166–71.

65. Magnetic Amplifier Sequential Circuits of the Asymmetrical Type, **Alan S. FitzGerald.** *Ibid.*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 685–91.

66. A New Static Magnetic Switching Circuit Using Combined "And" and "Or" Functions, **R. E. Morgan.** *AIEE CP 57-101*, 1957 (available on request).

67. Static Control in Automatic Warehousing, **L. L. Bosch, A. J. Fanthorp, J. W. Stuart.** *AIEE CP 57-189*, 1957 (available on request).

68. Static Switching Devices, **R. A. Mathias.** *Control Engineering*, New York, N. Y., May 1957, pp. 67–94.

69. Ferromagnetic Core Logic Circuitry and Its Application to Digital Computers, **A. J. Lincoln.** *Memorandum Report no. 911, PB 111954*, Office of Technical Services, U. S. Army, Washington, D. C., 1955.

70. Magnetic and Ferro-electric Computing Components, **V. L. Newhouse.** *Electronic Engineering*, London, England, May 1954, pp. 192–99.

71. Transcendental Function Analogue Computation with Magnetic Cores, **D. H. Schaefer, R. L. Van Allen.** *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 160–65; also *Report no. 4681*, Naval Research Laboratory, Washington, D. C.

72. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, **L. A. Finzi, R. A. Mathias.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 455–61.

73. A Magnetic Amplifier Switching Matrix, **D. Katz.** *Ibid.*, vol. 75, pt. I, May 1956, pp. 236–41.

74. A Single-Transistor Magnetic-Coupled Oscillator, **Kan Chen, A. J. Schiewe.** *Ibid.*, Sept. 1956, pp. 396–400.

75. A New Self-Excited Square-Wave Transistor Power Oscillator, **G. C. Uchrin, W. O. Taylor.** *Proceedings*, Institute of Radio Engineers, vol. 43, no. 1, Jan. 1955, p. 99.

76. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, **G. H. Royer.** *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

77. Operational Magnetic Amplifier with Audio-Frequency Transistor Power Supply, **R. O. Decker, Francis Gourash.** *Ibid.*, Sept. 1955, pp. 490–96.

78. Magnetic Amplifier Control of Switching Transistors, **H. William Collins.** *Ibid.*, vol. 75, pt. I, Nov. 1956, pp. 585–89.

79. Transistor Power Amplifiers with Switched Mode of Operation, **A. G. Milnes.** *Ibid.*, July 1956, pp. 368–72.

80. A Variable Frequency Magnetic-Coupled Multivibrator, **R. L. Van Allen.** *Ibid.*, vol. 74, pt. I, July 1955, pp. 356–61.

81. A Static Magnetic Transistor Analogue Divider, **D. H. Schaefer.** *Ibid.*, Nov. 1955, pp. 652–55.

82. Transductors and Magnetic Amplifiers (book), **A. G. Milnes.** MacMillan & Company, Ltd., London, England; St. Martin's Press, New York, N. Y., 1957.

◆

Discussion

H. J. McCreary (General Telephone Laboratories, Inc., Northlake, Ill.): This paper is a very timely review of the field of nonlinear magnetics. Since some of my pioneering work is cited in references 9 and 21 of the paper, perhaps some of my more recent work should be of interest to those pursuing this subject. In addition to the material cited are my U.S. patents: nos. 2,567,116, 2,602,153, 2,611,119, 2,611,120, and 2,623,205.

Of particular interest should be U.S. Patent no. 2,703,388, covering the magnetic cross valve used as an amplifier. Also of probable interest to those in this field is a recent paper which I presented at the 1957 National Electronics Conference, "The Magnetic Memory Switch or Madget Register Storage Device." This is a new method of data storage which has created considerable interest because it is more simple and direct than any other known means.

An Experimental Study of a Binary Code

W. W. PETERSON
ASSOCIATE MEMBER AIEE

ELIAS HAS DESCRIBED a coding system which consists essentially of iterating the Hamming code[1] and has obtained estimates of the probability of error after each iteration on the assumption that each set of check bits is used once and then discarded. It seemed that the first set of check bits could be reused profitably after some errors were corrected with the second set of check bits but the analysis is seriously complicated by the interdependence of errors. Therefore an experimental study of codes of this type was made using the IBM *704*. Matrices of 8 by 16, 16 by 32, and 32 by 64 bits with Hamming single-error correcting, double-error detecting codes in each column and each row, were used. Errors were introduced randomly, and the Hamming correction procedure was then applied alternately to rows and columns. The results of the experiment and some properties of the code pertinent to their discussion are described in this paper.

The Code

Information is assumed given in the form of rectangular matrices of binary digits. Columns of check digits are added to these arrays to make each row conform to the Hamming single-error correcting, double-error detecting code. Then rows of check digits are added to make each column conform to the Hamming code. It may be shown that the new rows conform automatically to the Hamming code so that a system of parity checks may be applied to every row and every column to correct and detect errors.

In the experiments, matrices of only three sizes were used: 8 by 16, (4-by-11 information bits), 16 by 32 (11-by-26 information bits), and 32 by 64 (26-by-57

Paper 58-38, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 15, 1957; made available for printing January 2, 1958.

W. W. Peterson is with the University of Florida, Gainesville, Fla.

The work described was carried out at the International Business Machines Corporation research center, Yorktown Heights, N. Y. This problem was suggested by Peter Elias, who contributed a number of valuable ideas, including the method for comparing codes. Jake Johnson wrote the most difficult part of the *704* code, that which actually covers the parity check and Hamming correction on a set of rows or columns.

TABLE I. ARRANGEMENT OF PARITY CHECKS FOR HAMMING CODE

CHECK COLUMN NO.	4	2	1	0
Column No.				
0				X
1			X	X
2		X		X
3		X	X	X
4	X			X
5	X		X	X
6	X	X		X
7	X	X	X	X

Un-checked	After Row Check	After Column Check	After 2nd Row Check
0011	0011	0010	0000
0010	0000	0000	0000
1100	1100	1100	1100
0100	0000	0010	0000
1010	1010	1010	1010
1000	0000	0000	0000
0010	0000	0000	0000
0110	0110	0110	0110

Fig. 1. Example of correction procedure for 4-by-8, all-zero matrix

(a)	(b)
1110	1111
1010	1010
1100	1100
1000	0000

Fig. 2. An uncorrectable pattern

information bits). If the rows or columns are numbered starting with zero, the check bit rows or columns are numbered, for convenience, 0, 1, 2, 4, 8, 16, and 32. The arrangement of parity checks for a row of length 8 is shown in Table I. The X's under a given check column number indicate which columns are included in that parity check, and the check bit is chosen to make the total number of ones in columns included in that parity check even. The arrangements for larger rows and columns are the obvious extensions of Table I.

Information with parity checks is assumed to be sent through a noisy channel so that certain bits in the array change. At the receiver end, first the parity of each row (corresponding to the parity check column 0) is checked. If that does check, there must either be no errors or an even number of errors. In either case nothing is done. If it does not check, then it is presumed that there is a single error in the row (although there may be any odd number of errors), and the correction is made by calculating the other parity checks and changing the only single bit that could cause the particular combination of parity check failures found to occur.

After this correction procedure is applied to rows, it is applied to columns, then alternately again to rows and columns until nothing more is gained.

The ability of the Hamming code, and hence of this code, to correct errors depends only upon the pattern of errors and not at all upon what the information bits happen to be. This is true of all group alphabets.[2] For simplicity, therefore, the codes were studied with all zeros in the information positions. It followed, then, that initially there had to be all zeros in the parity check positions as well. Thus any ones that remained after correction represented errors.

An example of the operation of the code is given in Fig. 1 for a 4-by-8, all-zero matrix. The errors introduced appear as ones in the unchecked matrix. When the error correction procedure is applied to rows, the second matrix results. The third matrix results after error correction is applied to columns, and the fourth after it is applied once more to rows.

This example illustrates two important facts: First, it demonstrates that after the first check on rows and on columns, additional correction may sometimes be gained by applying the checking procedure to rows again. Second, it shows that a point may be reached where the checking procedure described in this paper will produce no change in the matrix, since further operations on the fourth matrix will have no effect. Another type of stalemate in this checking procedure is illustrated in Fig. 2. A check on rows in matrix (*a*) will result in matrix (*b*), while a check on columns in (*b*) will result in (*a*) again.

This code has minimum distance 16,[3] since the smallest pattern of errors resulting in no parity check failures is a pattern including all the intersections of four rows and four columns. It therefore has the potential ability to correct any pattern of fewer than one half of 16, or 8, errors. The procedure described does not do this, however. Patterns of four errors left uncorrected by this procedure result in exactly two parity check failures in rows and two in columns. They may be corrected simply by changing the digits at the four intersections. Strangely, patterns of six errors may also be corrected by a similar procedure. They manifest themselves in the failure of parity checks in three rows and in three columns. Changing all bits of intersections will correct six errors and produce three errors, but the three new errors may be corrected by applying the correction procedure, at most, twice.

The error probability after the correction procedure has been applied once to rows may be estimated by the method used by Elias.[1] Because single errors are corrected, even multiple errors remain unchanged, and odd-multiple errors result in the addition of no more than one error:

$$Np_1 \leq \sum_{i=2,\ \text{even}}^{\leq N} ip(i) + \sum_{i=3,\ \text{odd}}^{\leq N} (i+1)p(i) \tag{1}$$

Here N is the number of digits in a row, p_1 is the probability of error per position after correction, and $p(i)$ is the probability of i errors in the N digits before correction. Where p_0 is the probability of error before correction

$$p(i) = \binom{N}{i} p_0^{\,i}(1-p_0)^{N-i} \tag{2}$$

This last expression may be put in the form

$$p_1 \leq \sum_{i=2}^{N} \binom{N}{i} \left(\frac{2i-2^{i-1}}{N}\right)(-p_0)^i \tag{3}$$

The first term is used by Elias as an upper bound on p_1. The terms alternate in sign, and therefore the sum of the first k terms for any odd k is an upper bound. Usually p_0 is small enough so that all but the first one or two terms are negligible. For small p_0 the equation gives a close estimate of probability of error: it may be shown that the correction procedure always converts a triple error into a quadruple error, and that therefore there is no approximation in the equation up to and including quadruple errors.

The error probability after correction has been applied to columns the first time may be estimated by this same procedure, since the probabilities of errors in the positions in any one column will be independent. In equation 3, p_1 should be used in place of p_0, and N should be taken as the number of digits per column. After correction has been applied on rows for the second time, however, the procedure breaks down because there is no longer statistical independence of the bit positions. Trying to account for the dependence appears altogether too complicated and so does trying to predict which error patterns are correctable and which are not. One way out of this impasse is an experimental evaluation.

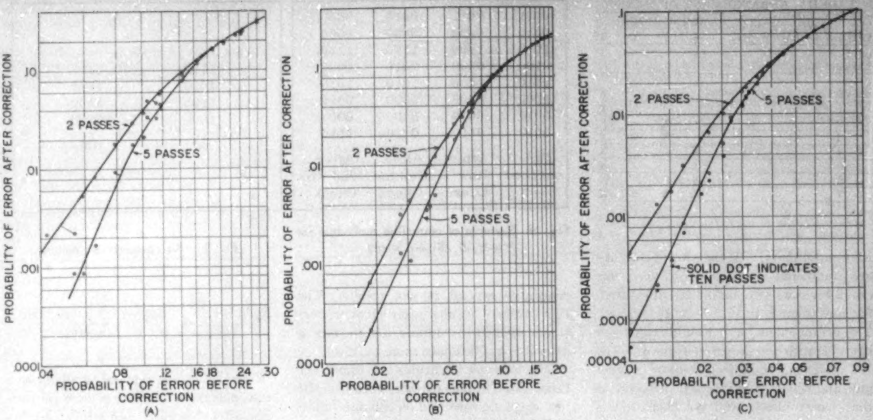


Fig. 3. Probability of error after correction

A—For an 8-by-16 matrix

B—For a 16-by-32 matrix

C—For a 32-by-64 matrix

The Experiment

For the experimental work, the IBM *704* subroutines were written to form a matrix with random errors, to count errors and print totals, to go through the Hamming correction procedure on rows or columns, and to print an entire matrix. The last subroutine was used only for checking the correction and error-counting routines on a number of examples.

Each 32-by-64 matrix consisted of a single bit in corresponding bit positions in each of 2,048 *704* words. Since each *704* word has 36 bits, the block of 2,048 words stored 36 matrices. The routine which formed the matrix first cleared the bits in all matrices to zeros, since the information bits were always made zeros in the experiments. Then for each bit in each matrix a pseudorandom number between zero and one was calculated by the currently popular method, and an error was introduced if the number was less than p_0. Successive powers of 5^{13} modulo 2^{35} were used.[4]

The method of storing the matrices enabled simultaneous calculation of parity checks for 36 matrices using the AND, OR, and COMPLEMENT instructions of the *704*. On the basis of the parity checks, the decision is made, for one matrix at a time and for one column or row at a time, as to whether and where correction should be made. For 32 by 64 matrices, the complete checking procedure applied once to rows or to columns of the 36 matrices required roughly 6 seconds.

The error-counting routine examined each bit in each matrix, one matrix at a time, and printed the total number of ones in each matrix (which was also the total number of errors, since the correct information and parity check bits were all zeros). It also printed the average frequency of error for all 36 matrices. A complete print-out required 10 seconds.

A typical experiment consisted of the formation of a matrix and printing of the number of errors before correction, followed by 10-correction cycles, each followed by a print-out, alternately, on rows and columns. A total of about 205 seconds was required for the 36 32-by-64 matrices.

Discussion of Results

The plots of average probability of error after decoding, Fig. 3, represent the most important results. For each point the abscissa is the percentage of positions (including check positions) with errors before correction, and the ordinate is the percentage of all positions (also including check positions) with errors after correction. In each case the upper curve gives error probability after the correction procedure is applied once to rows and once to columns. The lower curve corresponds to correction applied alternately to rows and columns a total of five times (three times to rows).

In Fig. 3(C), the graph for 32-by-64 matrices, a few points, shown as solid dots, are given for correction procedure applied ten times. For the 16-by-32 and 8-by-16 matrices, shown in (B) and (A), the points for ten passes were not plotted because they would lie so close to the points for five passes as only to cause confusion. It appears that there is little to be gained after the fifth pass, especially for the smaller matrices, where the difference was greater than 10% in only a few cases.

The improvement in probability gained by using five instead of two passes is greatest in the low probability range, amounting to as much as a factor of five. As the probability of error before correction increases, the difference gradually decreases until the curves meet at a point where the probability is about one third greater than the reciprocal of the matrix width. At that point, however, the probability of error after correction is approximately as great as the probability of error before correction, and the code would certainly not be used.

Equation 3 gives an upper bound for the probability of error after one pass, say, on rows. An upper bound on the probability of error after two passes may be determined by putting the result back into equation 3 with the appropriate value of N. Since the approximation in deriving equation 3 affected only sets of more than four errors in a row or column, the equation would be expected to yield an error probability that is in good agreement with experimental results after two passes. It does, in fact, give values which agree within 10% everywhere and 2% at most points with the upper curves in Fig. 3.

This equation cannot be used to estimate probability of error after three or more passes because the errors are no longer independent.

Elias has given upper and lower bounds for the probability of error for the best binary block code, given the total number of bits, the number of information bits, and the probability of error for the channel.[5] The values for the bounds are listed in Table II for codes corresponding to the points at which these matrix codes are most efficient in the sense defined in the next section. The optimum code corresponding to the 8-by-16 matrix code has error probability lower by only about a factor of five. For the larger matrices, however, the difference amounts to several orders of magnitude.

Some data on the frequency of error patterns of various sizes are given in Table III. For small probability of error before correction, patterns of four and six errors predominate. It is noted in the section "The Code" that all patterns of four or six errors may be corrected, though not by the Hamming correction on rows and columns.

Data on the number of passes required to bring a matrix to its final condition are given in Table IV for a few cases. In general, more passes are required when there are many errors than when there are few, and in only a relatively small number of cases were more than five correction passes required.

Two kinds of final conditions were noted: those with a constant and those with an oscillating number of errors. A constant number did not imply a constant pattern of error; the illustration in Fig. 2 is a case in point. The number of matrices in which the number of errors was observed to oscillate is indicated in Table IV for the cases presented. In general, the oscillation was more common when there were many errors. It was also noted that when there was oscillation, the smaller number of errors usually occurred after correction on rows, i.e., after an odd number of correction passes. This is quite reasonable because rows were shorter than columns and the Hamming check thus had more error-correcting ability on the rows.

TABLE II

BOUNDS ON THE ERROR PROBABILITY FOR OPTIMUM BINARY CODES

Total Bits	Information Bits	p *	Upper Bound	Lower Bound
128	44	0.1	1.5×10^{-2}	6.2×10^{-3}
512	286	0.05	6×10^{-5}	3×10^{-5}
2048	1482	0.025	2×10^{-9}	1×10^{-9}

* Probability of error per bit before decoding.

TABLE III

DISTRIBUTION OF NUMBER OF ERRORS
AFTER TEN CORRECTION PASSES

			Number of Errors in Pattern						
Matrix Size	p *	0	4	6	8	10	12	14	>14
8 by 16	0.0543	35	1	0	0	0	0	0	0
	0.1046	20	5	2	2	3	0	0	4
	0.1636	3	3	1	2	3	3	3	18
16 by 32	0.0183	35	1	0	0	0	0	0	0
	0.0306	33	2	0	1	0	0	0	0
	0.0499	17	5	3	2	1	1	2	5
	0.0667	3	3	2	2	2	4	2	18
32 by 64	0.0101	35	1	0	0	0	0	0	0
	0.0203	24	5	1	2	0	1	1	2
	0.0312	3	1	0	1	1	0	3	27

* Probability of error per bit before correction

TABLE IV

DISTRIBUTION OF NUMBER OF CORRECTION
PASSES BEFORE STEADY STATE IS REACHED

		Number of Passes								No. of cases of Oscillation
Matrix Size	p *	0	1	2	3	4	5	6	> 6	
8 by 16	0.0543	0	12	20	0	0	0	0	0	0
	0.1046	0	4	8	12	8	3	0	0	1
	0.1636	1	6	10	9	4	6	0	0	10
32 by 64	0.0101	0	5	22	7	1	1	0	0	0
	0.0203	0	0	1	8	9	10	5	3	1
	0.0312	0	0	3	9	7	7	4	6	17

* Probability of error per bit before correction.

A Method of Comparing Codes

A method of comparing two codes that differ in rate and error probability would be extremely useful. The following measure, suggested by Elias, seems quite appropriate, though without obvious theoretical basis. The quantity α, defined by the following equation, measures the efficiency of a binary code in some sense:

$$\frac{N_1}{N_2}=\left(\frac{1-H(p_2)}{1-H(p_1)}\right)^{\alpha} \qquad (4)$$

where N_1 is the number of binary symbols in a block before decoding, p_1 is the probability of error for any symbol before decoding, N_2 is the number of information symbols after decoding, p_2 is the probability of error for any symbol after decoding, and H is the entropy function.

Since information rate cannot be increased by a decoding operation[6]

$$N_1[1-H(p_1)] \geq N_2[(1-H(p_2)] \qquad (5)$$

and since $N_1 \geq N_2$, it follows that α is at least one. Furthermore, α equals one if, and only if, the equality in equation 5 holds, i.e., if no information is lost in decoding. It is clear that if N_1, N_2, and p_1 are the same for two different codes, then the one with the smaller p_2 is the better, and from equation 4 it may be seen that the better code has the smaller α.

Suppose that two codes had been used in cascade, the first to convert blocks of N_3 into respective blocks of N_2 symbols and the second to form a block of N_1 symbols from each N_2 block. Sending a block of N_1 symbols over a noisy channel then results in errors with probability p_1. Decoding to blocks of N_2 symbols results in probability of error p_2 and efficiency α_1. And further decoding to blocks of N_3

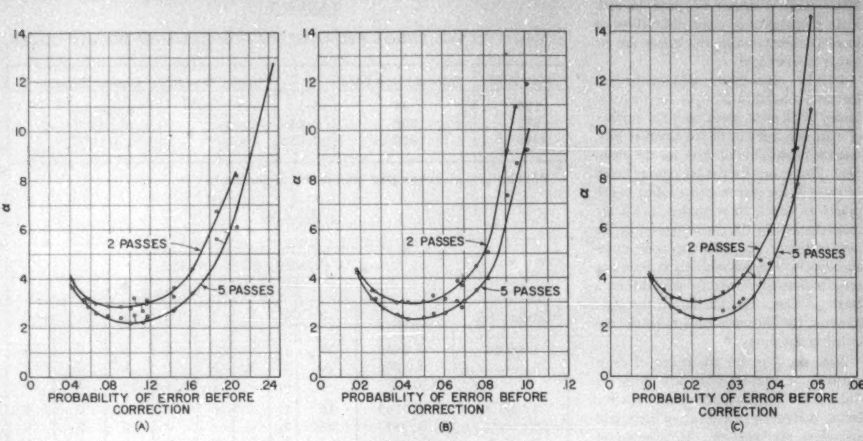


Fig. 4. Efficiency parameter α

A—For an 8-by-16 matrix

B—For a 16-by-32 matrix

C—For a 32-by-64 matrix

symbols results in error probability p_3 and efficiency α_2 for this second step. If $\alpha_1=\alpha_2$, the over-all α turns out to equal α_1 since

$$\frac{N_1}{N_3}=\frac{N_1}{N_2}\frac{N_2}{N_3}=\left[\frac{1-H(p_2)}{1-H(p_1)}\right]^{\alpha_1}\left[\frac{1-H(p_3)}{1-H(p_1)}\right]^{\alpha_1}$$
$$=\left[\frac{1-H(p_3)}{1-H(p_1)}\right]^{\alpha_1} \qquad (6)$$

If $\alpha_1\neq\alpha_2$, the over-all α may be found by an averaging process in which each α is weighted in proportion to the amount the corresponding code decreases the logarithm of the rate per binary symbol:

$$\alpha=\frac{\log N_1-\log N_3}{\log\,[1-H(p_3)]-\log\,[1-H(p_1)]}$$
$$=\left\{\frac{\log N_1-\log N_2}{\log\,[1-H(p_2)]-\log\,[1-H(p_1)]}\right\}\times$$
$$\left\{\frac{\log\,[1-H(p_2)]-\log\,[1-H(p_1)]}{\log\,[1-H(p_3)]-\log\,[1-H(p_1)]}\right\}+$$
$$\left\{\frac{\log N_2-\log N_3}{\log\,[1-H(p_3)]-\log\,[1-H(p_2)]}\right\}\times$$
$$\left\{\frac{\log\,[1-H(p_3)]-\log\,[1-H(p_2)]}{\log\,[1-H(p_3)]-\log\,[1-H+(p_1)]}\right\}$$
$$=\alpha_1\lambda+\alpha_2(1-\lambda) \qquad (7)$$

where

$$\lambda=\frac{\log\,[1-H(p_2)]-\log\,[1-H(p_1)]}{\log\,[1-H(p_3)]-\log\,[1-H(p_1)]}$$

Thus for cascaded codes the efficiency parameter α may be calculated in a convenient manner from the α's for the individual codes.

The values of α for the experiments run on the IBM *704* are plotted in Fig. 4. It is a surprising fact that the minimum (best) α is very nearly the same for all three codes and occurs at an input probability almost inversely proportional to the width of the matrix. The curves are remarkably similar for the three matrix sizes.

Conclusions

Although the code described in this paper falls far short of being optimum for its length, the fact is that no one yet knows how to construct optimum codes except for isolated cases. The code does, however, permit correction procedures that are relatively easy to implement with modern digital techniques, and for that reason should be of more than theoretical interest. By the same token, modern digital computers may prove an important tool for the information theorist in so far as they facilitate experimentation in an area in which little experimental work has hitherto been done.

References

1. Error Free Coding, P. Elias. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N.Y., PGIT-4, Sept. 1954, pp. 29–37.

2. A Class of Signaling Alphabets, David Slepian. *Bell System Technical Journal*, New York, N. Y., vol. 35, Jan. 1956, pp. 203–34.

3. Error Detecting and Error Correcting Codes, R. W. Hamming. *Ibid.*, vol. 29, 1950, pp. 147–60.

4. Generation and Testing of Pseudo-Random Numbers, O. Taussky, J. Todd. "Symposium on Monte Carlo Methods," H. A. Meyer, editor. John Wiley & Sons, Inc., New York, N. Y., 1956, pp. 15–28.

5. Coding for Noisy Channels, P. Elias. *IRE Convention Record*, Institute of Radio Engineers, pt. IV, Mar. 1955, pp. 37–46.

6. The Mathematical Theory of Communication (book), Shannon, Weaver. University of Illinois Press, Urbana, Ill., 1949, p. 27.

Magnetic Saturable-Core Timing Device

J. L. LOWRANCE
ASSOCIATE MEMBER AIEE

ACCURATE TIMING DEVICES are fundamentally important in both industial and military control applications. This paper deals with electric timing apparatus in which the timing sense is derived from the volt-second time integral required to saturate a saturable magnetic core. Discrete time periods may be generated on the basis of this principle. In this paper, fundamental design equations are presented and basic circuitry is described. Specific applications are discussed which make use of transistors in the switching mode. Transistors afford convenient means for operating various electric circuits as a function of the time period defined by the controlled saturation of the magnetic cores. The transistors contribute no appreciable time delay in the circuit. The problems associated with environmental variations in transistors and cores are discussed, and possible methods of compensation are presented.

The changes in the magnetic flux level of a magnetic reactor may be related to the volt-time integral of the voltage induced in the windings of the reactor. For fixed values of the other parameters, the time period required to saturate the core is a constant.[1]

$$\Delta T=\frac{2AB_mN}{e}10^{-8} \qquad (1)$$

Therefore a definite time period may be generated using the saturable magnetic core. The accuracy of this time period is is a function of accuracy to which the other parameters are held constant, and the accuracy of sensing the time at which the core becomes saturated. The saturation of the magnetic core is accompanied by a change in the inductance of the saturable reactor. If the core material is of the square-loop variety, this change will be very pronounced. A radical change occurs in the impedance of the electrical circuit associated with the saturable reactor because of this inductance change. The change in impedance can be used to trigger another circuit at the expiration of the time period, when the core saturates.

Type of Core Material

The type of core material used in this application is governed by consideration of several factors. As just mentioned, it is desirable to use a square-loop core material to insure a definite and discrete enunciation in the circuit when the core saturates. The time period is directly related to the cross-sectional area of the core. Therefore it is advantageous to make the saturation flux density of the core as large as possible in order to minimize the size and weight of the saturable reactor.

The magnetic material that most nearly fulfills these qualifications is the 50% nickel, 50% iron alloy. Some commercial names for this alloy are Orthonol, Deltamax, and Orthonik. The coercive force is about 0.06 oersted, and the residual flux may be made between 12,000 and 15,000 gausses.[2] The squareness of the hysteresis loop is quite pronounced in this material. Therefore the core provides the desired sharp change from a high to a very low inductance in the circuit. Fig. 1 shows the d-c hysteresis-loop for Orthonol.

It should be mentioned that cores fabricated as tape-wound toroids afford the best performance. The effective air gap is quite small, permitting the ideal characteristics of the core material to be approached. Since the core is operated on its d-c hysteresis loop, it is not essential that the core have thin laminations. The most desirable core material is a tape thin enough to provide a small effective air gap, yet thick enough to maximize the stacking factor.

Description of Circuitry

The volt-time integral characteristic of the magnetic saturable reactor may be used in various ways to generate a time period or delay interval. Three possible modes of operation will be discussed.

SERIES SATURABLE REACTOR CIRCUIT

For an industance in series with a resistance, the expression for the current flow is

$$E=L\frac{di}{dt}+Ri \qquad (2)$$

If $L>>R$, the current is primarily a function of the magnetizing current of the inductance. This current is

$$I=\frac{Hl_m}{0.4\pi N} \qquad (3)$$

Fig. 2(A) shows a curve of current versus time for a series saturable reactor-resistance circuit. When the core saturates, the current becomes equal to the applied voltage divided by the resistance of the circuit. Putting the saturable reactor in series with the input circuit of a transistor, this change in current may be used to control the output circuit of the transistor. The combined circuit will act essentially as a switch that is shunted by a high resistance in series with a low-impedance load.[3] A grounded emitter transistor circuit is shown in Fig. 3(A). The saturable reactor is initially set to negative saturation by the reset winding N_3. When the switch is thrown to the "fire" position, voltage is applied to the emitter-collector circuit and the Zener diode voltage-regulating circuit.[4] The voltage-regulating

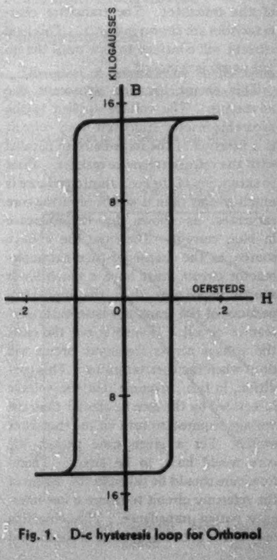


Fig. 1. D-c hysteresis loop for Orthonol

Paper 58-120, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 3, 1957; made available for printing November 8, 1957.

J. L. LOWRANCE is with the Bendix Aviation Corporation, South Bend, Ind.

The author wishes to thank J. Frick, L. Peterson, and L. White, of Bendix Products Division, for their assistance in providing experimental models and data used in this paper; and to acknowledge the support of WADC *Contract No. AF 33(600)-303 18*.

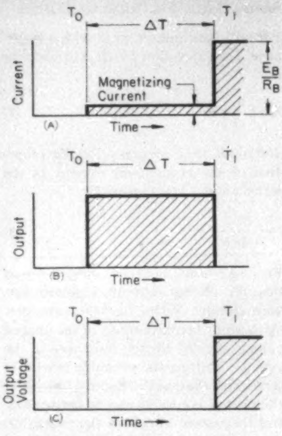


Fig. 2. Three modes of timer operation

circuit, in turn, supplies an essentially constant voltage to the transistor-base–saturable-reactor circuit. This constant voltage is integrated by the saturable reactor until it saturates at a predetermined time T_1. At time T_1, the base current increases and a current flows in the load circuit which is equal to the product of the base current and the current gain of the transistor. The transistor characteristics are shown in Fig. 4. This load current will continue to flow until the input power is removed.

This circuit has one significant disadvantage. The voltage applied to the saturable-reactor–transistor-base circuit is a function of the impedance in parallel with the voltage reference resistor. Prior to saturation of the core, this impedance is much greater than it will be after the core saturates, as shown by the increase in base current. Therefore the voltage source to the transistor-base–saturable-reactor circuit must have a much lower source impedance than the input impedance of the transistor-base–saturable-reactor circuit. If such is not the case, the voltage across the input circuit will drop when the core saturates. This condition, in turn, requires that the voltage integrated by the core be greater than the voltage required to turn on the transistor switch. For a given time period, the core would have to be larger. Therefore, care should be taken in the design of the reference circuit to insure a low reference output impedance. This objective may not be practical because of the limited voltage ratings of presently available "Zener reference diodes."

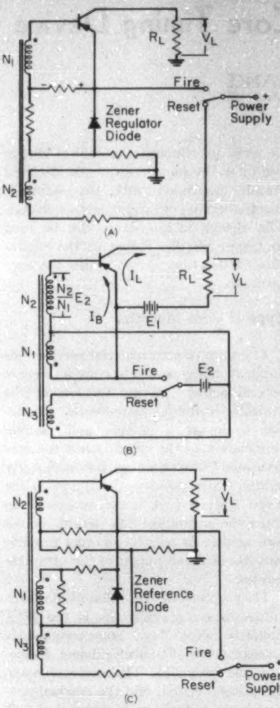


Fig. 3. Time delay circuit

A—See Fig. 2(A)
B—See Fig. 2(C)
C—With voltage regulation

Another disadvantage of this series saturable reactor circuit is an undesirable effect of the magnetization current that flows prior to the saturation of the core. This magnetization current also flows in the base circuit of the transistor, and consequently a correspondingly larger current will flow in the load circuit. This magnetizing current may be bypassed by shunting the base-emitter circuit of the transistor, with a resistor. This resistor would also be in series with the saturable reactor. The threshold voltage of the transistor would prevent current flow in the base-emitter circuit during the magnetization period.

A saturating transformer may be used to generate a pulse of voltage for a definite time interval if the impressed voltage is held constant. This pulse of voltage may be used either to turn on a circuit or to keep a circuit turned off for a definite period. Fig. 2(B) shows the waveform of the voltage across the low-voltage winding of the transformer. This phenomenon is used in the circuit shown in Fig. 3(B) to generate a time delay in the transistor circuit.

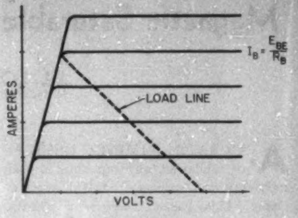


Fig. 4. Typical transistor grounded emitter characteristics

The core is set to one end of the hysteresis loop by the voltage E_2 across winding N_3. At time T_0 the switch is thrown to the "fire" position. The voltage

$$E_2\left[\frac{N_3}{N_1}-1\right] N_3>N_1 \tag{4}$$

appears in the base circuit. The net voltage appearing across the base-emitter circuit is in the direction to keep the transistor turned off. At the time that the core saturates, the voltage across the base-emitter circuit becomes equal to E_2 and the transistor will turn on. The load current will then flow in the output circuit as shown in Fig. 2(C). It is advantageous to make E_2 large enough to saturate the transistor for a given load R_L. This allows the voltage

$$E_1-I_LR_{ce} \tag{5}$$

to appear across the load, and the transistor is thereby operated in its most efficient mode. The saturated junction resistance of the transistor R_{ce}, is quite small for most germanium power transistors. The small value of this resistance allows the transistor to be operated at ambient temperatures near the maximum storage temperature. The transistor will remain on until power is removed from the base circuit. The core may be reset and the timing operation may then be repeated. The reset circuit should be designed so that the voltage reflected into the base-emitter circuit is less than the threshold voltage of the base-emitter circuit of the transistor.

The time period is directly proportional to the voltage absorbed by the core. Therefore a practical circuit application must incorporate some means of regulating the voltage across the winding of the transformer to be sure that the time period will be constant. Fig. 3(C) shows a circuit that uses a Zener diode as a voltage regulating reference device. The voltage across the diode is essentially constant after a certain Zener voltage E_{zr} is reached. This Zener diode regulates the voltage across the timer control winding and the voltage appearing in the base-emitter bias circuit. For most commercially available reference units with zero temperature coefficient, the reference voltage must be attenuated with a resistance network, because of the magnitude of the Zener voltage. Prior to saturation of the core, the impedance of winding N_1 is large compared to its shunting resistor. Therefore the diode voltage will divide according to the resistance ratio shunting the diode. The magnetizing current will flow through winding N_1. This condition will have to be considered if this magnetizing current is large compared to the total current in the attenuating circuit.

Fig. 5 shows a circuit similar to the circuit shown in Fig. 3(C). A 2-stage switching amplifier has been added to the basic timing circuit. When the switch is thrown to the "fire" position, the following events occur:

1. The base-emitter bias circuit of P_1 is energized.
2. The timing voltage across N_1 is applied.
3. Current flows through P_2 in to the load of P_1, turning P_2 on and keeping P_3 turned off. E_L is zero.

At the time that the core becomes saturated, the voltage induced in winding N_2 crops to zero. The base-emitter current of P_1 turns P_1 on. This shunts the base-emitter circuit of P_2. The base potential of P_2 becomes essentially equal to the emitter potential of P_2 and P_2 is therefore turned off. The base of P_3 is no longer at the same potential as its emitter, so it is turned on. Voltage then appears across the load R_L.

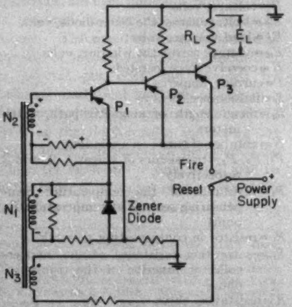


Fig. 5. Three-stage transistor switch and timer circuit

As mentioned previously, the saturable transformer may be used to generate a pulse of voltage of a definite time length. A circuit that performs this function is shown in Fig. 6. Here the voltage induced in the emitter-base circuit is of the polarity to turn on the transistor rather than to turn it off. Therefore the transistor is biased fully on for the period of time necessary for the core to saturate. At saturation, the bias voltage reduces to zero and the transistor is again turned off. This circuit has the disadvantage of drawing power from the control source, which introduces problems of temperature compensation for maintaining a constant reference voltage for the winding. The winding impedance will change as the base-emitter resistance of the transistor varies with temperature, and as the winding resistances of the high-voltage and low-voltage transformer windings vary with temperature. If it is necessary that the circuit be conducting for this period, the stages of transistor amplification may be arranged to turn off rather than on when the time period is being actuated, thereby alleviating the temperature compensation problem.

Magnetic-Field Effect

The time period required to saturate the core is a direct function of the saturation flux density of the core material. It has been pointed out in several current papers that the saturation flux density of magnetic material is a function of the strength of the magnetic field imposed perpendicular to the field of the excitation winding. The net effect is a reduction of the saturation flux density as the perpendicular field strength is increased. The mathematical analysis of this phenomenon may be found in reference 5.

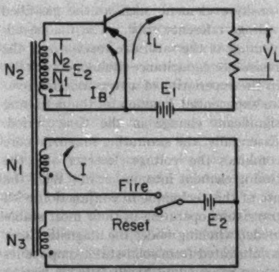


Fig. 6. Pulse output timer circuit

It may be expected that an external magnetic field will have an effect upon the time period of a magnetic timer; this has been found to be a valid assumption. The time required to saturate the core at a given voltage is shortened by the influence of this field. Fig. 7 is a typical curve showing the change in time period as a function of field intensity. Curve *A* shows the effect of a d-c field. Curve *B* shows the effect of the same field intensity with magnetic shielding of cold-rolled steel around the saturable reactor assembly. Curve *C* shows the effect of an a-c field on the same shielded assembly. The choice of location of the timer assembly in the magnetic field is important. The greatest effect is caused by fields parallel to the axis of the toroid. It is felt that the field intensity necessary to affect the time period appreciably is much larger than would be encountered in most applications. If the field were predictable it could be compensated for in the design of the timer.

This field effect is interesting in view of the possible control applications involving this phenomenon. It would be possible to vary the time period as a function of some parameter which controlled the field intensity around the core.

Temperature Effect

The effect of temperature on the time period may be divided into two categories:

1. The change in saturation flux of the core.

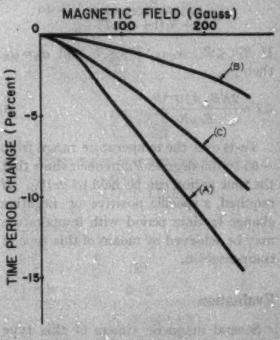


Fig. 7. External magnetic field intensity period

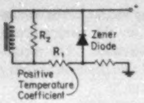


Fig. 8. Temperature compensating circuit

2. The effect of temperature on other circuit components. To some extent, these differences may be made to compensate for each other.

The change in time period is inversely proportional to temperature, causing the temperature coefficient of saturation flux density to be negative. The slope of the curve is not the actual temperature coefficient of flux density, because the increase in cross-sectional area of the core caused by an increase in temperature would tend to compensate for the corresponding decrease in flux density. Nonetheless, the slope of the curve is particularly useful for compensation purposes, since the net effect of the change in B_mA determines the change in the time period.

One way to compensate for the change in time period with temperature is to vary the voltage applied to the core so as to cancel the effect of the change in B_mA. A circuit that achieves this compensation is shown in Fig. 8. The voltage appearing across the winding is

$$E_w = E_{CR}\frac{R_2}{R_2+R_1(1+\alpha_R\Delta T)} \tag{6}$$

This method of compensation is difficult to achieve because of the nonlinearity of the temperature coefficients of both the magnetic material and the compensating resistors. If the temperature coefficients are assumed to be constant, the equation for compensating the time period is

$$\Delta T = \frac{2NB_mA(1-\alpha_B\Delta T)}{E_{CR}\dfrac{R_2}{R_2+R_1(1+\alpha_R\Delta T)}}10^{-8} \tag{7}$$

If $R_2 << R_1$, $\alpha_B\alpha_R\Delta T^2 << 1$, and $\alpha_B=\alpha_R$, then

$$\Delta T \approx \frac{2NB_mA10^{-8}R_1}{E_{CR}R_2} \tag{8}$$

Tests over the temperature range from −65 to 165 degrees Fahrenheit show that the time period can be held to ±1%. If required, a specific positive or negative change in time period with temperature may be achieved by means of this type of compensation.

Evaluation

Several magnetic timers of this type have been fabricated for use in rocket engine controls. Fig. 9 shows a complete unit that provides two time periods, 0.22 and 1.30 seconds. This unit is capable of delivering 5 amperes to an inductive load. The volume is approximately 20 cubic inches and the weight is 1.5 pounds. The following components were used:

Transistors, Honeywell *H-5*, and General Electric *2N187*.

Reference diode, Hoffman semiconductor *1N429*.

Saturable cores, Magnetics, Inc., *51190-4A*, (1.30 seconds); and Magnetics, Inc., *51181-2A*, (0.22 second).

Input voltage, 17 to 29 volts d-c.

The circuitry used is similar to that of Fig. 5 with innovations for temperature compensation and switching of the two time periods in the output circuit.

At a vibration level of approximately 40,000 the 1.30-second time period decreases significantly. This effect is presumably caused by the change in the magnetic properties of the core at this high vibration level because of the mechanical stresses set up in the core material. The shorter time period exhibited essentially stable characteristics over this vibration range. The greater mass of the magnetic core used in the 1.30-second circuit afforded a more significant damping problem than the small core. It was extremely difficult to obtain quantitative data on this phenomenon because of the the difficulty, under these dynamic conditions, of measuring the saturation flux of the core accurately.

Conclusions

The volt-second integral associated with the square hysteresis loop of certain magnetic core materials is a useful and versatile basis upon which to design circuitry for generating discrete time periods. In the past, this field of application was dominated by the use of the time constant of the resistance-capacitance network in conjunction with some voltage-sensitive element such as the gas-filled voltage reference tube. The unique advantage of the saturable reactor over the resistance-capacitance timing network lies in the superiority of a step function over an exponential function as the actuating significant change in the time period. Essentially, the saturating magnetic core combines the voltage detector and the timing element into one unit. With the use of this technique in conjunction with transistors operating in their most stable mode (switching mode), the magnetic timer is fabricated from solid-state components into a rugged and highly reliable timing device. This construction makes the magnetic timer a complementary system component suitable for use with other magnetic and semiconductor control apparatus.

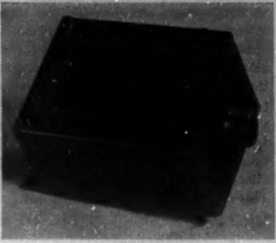


Fig. 9. Fabricated timer unit

The accuracy of device is a function of the compensation provided in the circuitry associated with the saturable core. Temperature compensation may be used to generate timing functions with positive, negative, or zero slope. The size of the timing device is directly related to the desired time period. Therefore the time period is limited, for practical reasons, to less than approximately 10 seconds. Novel circuitry may extend this practical limitation.

Appendix. Nomenclature

A = area of the cross section of the core, square centimeters
α_B = temperature coefficient of saturation flux density, gausses per gauss per C (degrees centigrade)
α_R = temperature coefficient of the resistor, ohms per ohm per C
B_m = saturated flux density of the core material, gausses per square centimeter
e = instantaneous voltage, volts
E = voltage, volts
E_{cr} = voltage across the Zener diode, volts
E_L = load voltage, volts
E_w = voltage across the winding, volts
H = coercive force, oersteds
I = current, amperes
L = inductance, henrys
l_m = mean length of magnetic path, centimeters
N = number of turns on a winding
P_1, P_2, P_3 = transistors of stages 1, 2, and 3 respectively
R_1 = resistance of the temperature compensating resistor at temperature T_0, ohms
R_2 = resistor in compensating network
R_{ce} = saturation resistance of the emitter-collector junction of the transistor, ohms
R = resistance, ohms
t = time, seconds
T_0 = reference temperature, C
ϕ = flux, gausses

References

1. PRINCIPLES OF ELECTRICAL ENGINEERING (book), William H. Timbie, Vannevar Bush. John Wiley & Sons, Inc., New York, N. Y., 1951, p. 301.

2. FERROMAGNETISM (book), Richard M. Bozorth. D. Van Nostrand Company Inc., Princeton, N. J., 1951, p. 133.

3. JUNCTION TRANSISTORS USED AS SWITCHES, R. L. Bright. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

4. THE SUITABILITY OF THE SILICON ALLOY JUNCTION DIODE AS A REFERENCE STANDARD IN REGULATED METALLIC RECTIFIER CIRCUITS, D. H. Smith. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 645–51.

5. SUPERPOSED MAGNETIC FIELDS IN MATERIALS WITH RECTANGULAR HYSTERESIS LOOPS, C. B. Wakeman, F. J. Beck. *Ibid.*, vol. 75, pt. I, Nov. 1956, pp. 562–69.

Accurate Determination of Capacitance by Conjoining Analytical and Analog Techniques

J. D. HORGAN
ASSOCIATE MEMBER AIEE

J. A. PESAVENTO
ASSOCIATE MEMBER AIEE

THE PRINCIPAL OBJECTIVE of the work described in this paper is to develop a new method for accurately determining the capacitance associated with 2-dimensional composite regions. The term capacitance refers to the ordinary electric capacitance associated with two charged conductors. However, the method is also directly applicable to the determination of the permeance of magnetic fields, to the shape factor of thermal fields, and to other analogous situations. The term "accurately determining" indicates determination of the capacitance with errors in the order of one part in 1,000. While such accuracy has been claimed for certain electrolytic tank analog methods[1] it is usually possible only after careful attention is given to each source of error. Such errors can arise in both the construction and the measurements associated with the analog solution. In the present method electrically conducting paper models, much easier to construct, are used and only ordinary care is required in the measurement techniques. The usual errors encountered in the use of such paper models, amounting to several per cent, are avoided by utilizing analytical methods to obtain a first approximation amounting to perhaps 90% of the final result, and using the paper analogs to determine the other 10%. Thus, the error in the final result will be in the order of a few per cent of, say, 10% or in the order of a few tenths of 1%.

"Two-dimensional region" refers to a region of electric field in the vicinity of two cylindrical (not necessarily circular) conductors or, in other words, a region of which the electric field is independent of one of the co-ordinate variables. A composite region is one which may be divided by means of thin insulating or conducting partitions into two or more simple regions. A simple region is one in which the electrical potential function can be exactly determined by some known analytical method, for example, by use of the Schwarz-Christoffel transformation. Several examples of composite regions are shown in Figs. 1 to 4.

Fig. 1 depicts two parallel rectangular conductors. The quadrant of the 2-dimensional composite region associated with the conductors may be divided by insulating partitions, as indicated by the dashed lines, into three simple regions, each of which may be analyzed for the potential function, with the use of the Schwarz-Christoffel transformation. The strip transmission line of Fig. 2 is associated with a composite region which may be similarly divided. The capacitance per unit length for this configuration has been determined by a series-matching method.[2] The coupled-strip transmission line of Fig. 3 is associated with a more complicated composite region which divides into three simple regions as indicated. The capacitance per unit length for this configuration has been worked out by an analytical method closely related to the method described in this paper.[3] Fig. 4 depicts the problem of the telephone relay, different from the preceding examples because it is a magnetic field problem and because the dividing partition is conducting rather than insulating. This problem has been investigated with the use of an electrostatic analog[4] and by relaxation methods.[5]

These examples serve to illustrate the type of problem for which the new method has been developed. With the use of the present technique, the capacitance of each of these configurations may be determined with errors in the order of 0.1%. In addition, work with this technique yields considerable insight into the magnitudes of the errors associated with the usual first approximations to capacitance obtained by simple partitioning of the field regions.

The Analysis

Explanation of the method can most easily be accomplished by reference to a specific problem. The strip transmission line serves as an excellent example because

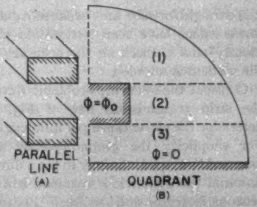


Fig. 1. Transmission line comprised of two parallel rectangular conductors

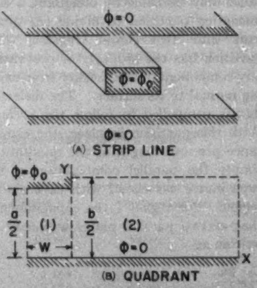
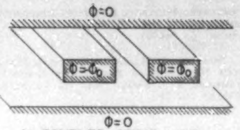


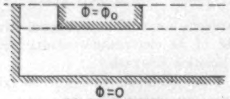
Fig. 2. Double-ground-plane strip transmission line

Paper 58-65, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 22, 1957.

J. D. HORGAN and J. A. PESAVENTO are with Marquette University, Milwaukee, Wis.

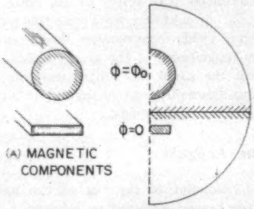


(A) COUPLED-STRIP LINE



(B) QUADRANT

Fig. 3. Coupled strip transmission line



(A) MAGNETIC COMPONENTS

(B) SEMI-PLANE

Fig. 4. Schematic of telephone relay

it is straightforward and because capacitance values have been worked out previously[3] and a check on the accuracy of the method is afforded.

Consider the following problem: Given the strip transmission line of Fig. 2, determine the capacitance per unit length. For simplicity the ground planes are assumed to extend infinitely in the x direction and the dielectric is assumed homogeneous, isotropic, and loss-free. Because of geometrical symmetry, only the quadrant of Fig. 2 need be analyzed.

A first approximation to the capacitance may be made by imagining a thin insulating partition to divide the quadrant into two simple regions. This partition has the property of preventing any component of flux density from existing normal to its surface. The dielectric flux is compelled to follow the surface. With this partition in place, the capacitance per unit length of the quadrant is simply the parallel combination of the capacitance associated with each simple region. For region 1 the capacitance is that of two parallel plates and may be written as

$$\frac{C_1}{\epsilon}=\frac{2w}{a} \tag{1}$$

where C_1 is the capacitance per unit length of region 1 and ϵ is the permittivity of the composite region. For region 2 the capacitance per unit length is derived in reference 3 with the Schwarz-Christoffel transformation. The result is

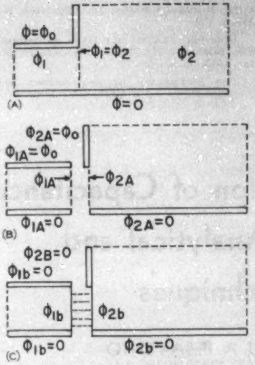


Fig. 5. The compensating fields

$$\frac{C_2}{\epsilon}=\frac{K(k)}{K(k')} \tag{2}$$

where C_2 is the capacitance per unit length of region 2, $K(k)$ and $K(k')$ are the complete elliptic integrals of the first kind, as tabulated, for example, in Dwight[6] and the arguments k and k' are related to the dimensions by

$$k=\cos\left(\frac{\pi a}{2b}\right) \tag{3}$$

$$k'=\sin\left(\frac{\pi a}{2b}\right)$$

The total capacitance per unit length for the strip transmission of Fig. 2 may thus be written

$$\frac{C}{\epsilon}=\frac{8w}{a}+\frac{4K(k)}{K(k')}+\frac{\Delta C}{\epsilon} \tag{4}$$

The first two terms on the right side of equation 4 represent the first approximation to the capacitance and $\Delta C/\epsilon$ represents an increment to be added to account for the presence of the insulating partition. In a previously described method[3] this increment has been determined analytically. In the present method it is determined from measurements on an electroconducting analog.

The Compensating Field

The analog to be constructed will represent a compensating field which accounts for the presence of the insulating partition. The nature of the construction may be understood by referring to Fig. 5, where (A) depicts the composite region which is being investigated, together with the appropriate boundary conditions, and (B) shows the two fields which have been analyzed to obtain the first approximation for capacitance. These two fields taken together satisfy all the boundary conditions of the original region except those at the partition or cut. Fig. 5(C) shows two field configurations which may be superposed with those in (B). When this superposition is made, the resultant configuration is identical with the original, providing that, at the partition, the total potential is continuous and the total of the normal components of flux density is continuous, i.e.,

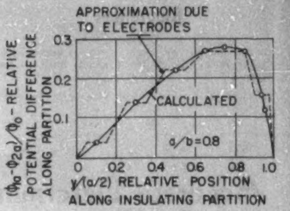


Fig. 6. Difference of field potentials $\phi_{1a}-\phi_{2a}$ at the partition

$$\phi_{1a}+\phi_{1b}=\phi_{2a}+\phi_{2b} \tag{5}$$

$$\frac{\delta\phi_{1a}}{\delta x}+\frac{\delta\phi_{1b}}{\delta x}=\frac{\delta\phi_{2a}}{\delta x}+\frac{\delta\phi_{2b}}{\delta x} \tag{6}$$

at the surface. Because of the insulating partition, the first term on either side of equation 6 is zero, so that

$$\frac{\delta\phi_{1b}}{\delta x}=\frac{\delta\phi_{2b}}{\delta x} \tag{7}$$

at the surface. For convenience, equation 5 is written as

$$\phi_{1b}-\phi_{2b}=-(\phi_{1a}-\phi_{2a}) \tag{8}$$

Inspection of equations 7 and 8 and of Fig. 5(C) will indicate that the boundary conditions required on the compensating field are as follows: 1. Segments of the boundaries corresponding to the conductors in the original problem must be held at zero potential. 2. Segments of the boundaries corresponding to flux lines in the original problem must be flux lines in the corresponding fields. 3. Along the surface of the partition, the normal components of flux densities in the compensating fields must be equal. 4. Along the surface of the partition, the difference of the potentials of the two compensating fields must equal the negative of the difference of the

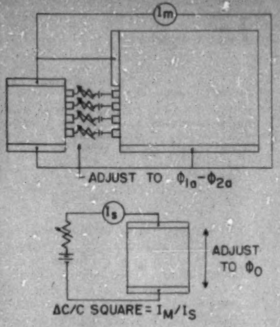


Fig. 7. Electroconducting analog

potentials of the two fields of Fig. 5(B). This difference is calculable. Details of the calculations are shown in reference 3 and typical results given in Fig. 6.

The Analog

With the complete boundary conditions for the compensating fields stipulated, it should be possible to construct the corresponding electroconducting analog. The diagram of such an analog, shown in Fig. 7, is constructed of Teledeltos paper[7] on which boundaries have been placed using silver-conducting paint. All four segments corresponding to the conductors in the original problem are connected together to satisfy the first boundary condition, given in the foregoing. Since current in the analog represents electric flux in the actual field, and since total flux equals total charge at the conductors, the current I_m is proportional to the charge ΔQ on the conductors of the compensating fields of Fig. 5(C). The increment in capacitance due to the compensating fields is thus given by

$$\Delta C=\frac{\Delta Q}{\phi_0}=\frac{\epsilon I_m}{\sigma\phi_0} \tag{9}$$

where σ represents the unit conductance of the electroconducting paper. A convenient means of eliminating σ from consideration is to apply the potential ϕ_0 to a square segment of the paper as indicated in Fig. 7 and read the current I_s. Then

$$\sigma=\frac{I_s}{\phi_0} \tag{10}$$

and the increment in capacitance may be obtained from

$$\frac{\Delta C}{\epsilon}=\frac{I_m}{I_s} \tag{11}$$

The third boundary condition requires that a current flow path be provided between the surfaces corresponding to the partition. For practical reasons, the desired single path is approximated by a finite number of paths in parallel as indicated in Fig. 7. The fourth boundary condition is fulfilled approximately by adjusting the differences in potentials in the current paths according to the curves of Fig. 6, in which figure the stepped curve shows how this potential difference occurs in the analog.

Measurements and Results

To check the accuracy of the method, analog models were constructed for the ratio a/b in Fig. 2, equal to 0.2, 0.4, 0.6, and 0.8 and the ratio w/b approaching infinity. On each model, the boundary potentials were adjusted according to Fig. 6, after selecting a convenient reference potential, ϕ_0. Then the currents I_m and I_s were recorded and the increment in capacitance calculated from equation 11. Results are shown in Table I, in comparison with results of reference 2. For a fair appraisal, per-cent errors are based on the usual fringe capacitance C_f, which, by inspection of equation 4 is given by

$$\frac{C_f}{\epsilon}=\frac{4K(k)}{K(k')}+\frac{\Delta C}{\epsilon} \tag{12}$$

It should be pointed out, however, that when the per-cent errors are expressed relative to the total capacitance, the per-cent error would be less than that given in Table I. It is also significant that these results were obtained with the rather crude 4-step approximation to the potential difference curves of Fig. 6. Had more steps been constructed, even more precise results could be expected. To obtain a comparison of the results of this technique with the more conventional analog method of predicting capacitance-conducting, paper models of the complete field of Fig. 5(A) were constructed, and the capacitance determined by comparing the model resistance with the resistance of a square of conducting paper in the bridge circuit of Fig. 8. The results of this measurement are shown in Table II. Since the conventional method measures total capacitance rather than fringe capacitance, comparisons of total capacitance are made in this table. If fringe capacitances were to be compared, the advantage of the new method would be even more apparent.

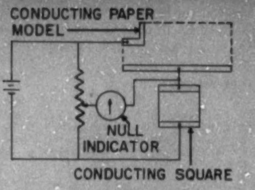


Fig. 8. Conventional analog method for capacitance determination

Conclusions

The usual analytical methods for determining the capacitances associated with 2-dimensional composite regions are generally inadequate to handle any but the simplest of these problems. The usual analog methods are either limited in accuracy, or require extremely precise experimental techniques. The method presented here represents a combination of analytical methods and analog tech-

Table I. Fringe Capacitance for Strip Transmission Line of Fig. 2

Fringe Capacitance C_f/ϵ

a/b	By Current Method	By Method From Begovich[3]	Per Cent Error
0.2	6.622	6.630	−0.12
0.4	4.804	4.810	−0.12
0.6	3.662	3.673	−0.35
0.8	2.770	2.765	+0.18

Table II. Comparison of Conventional Analog Method and New Method

Total Capacitance C/ϵ

			By New Method		By Conventional Method	
a/b	2w/a	Reference 2	Value	Per Cent Error	Value	Per Cent Error
0.2	20	86.630	86.622	−0.009	84.84	− 2.11
0.4	10	44.810	44.804	−0.012	43.40	− 3.25
0.6	7.5	33.673	33.662	−0.033	31.12	− 8.2
0.8	5	22.765	22.770	+0.022	20.24	−12.5

niques which retains the high accuracy attainable with analysis and the simplicity of the ordinary analog approach.

References

1. The Accurate Mapping of Electric Fields in an Electric Tank, **K. F. Sander, J. G. Yates.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. II, no. 74, 1953, pp. 167–75.

2. Capacity and Characteristic Impedance of Strip Transmission Lines with Rectangular Inner Conductors, **N. A. Begovich.** *Transactions*, Professional Group on Microwave Theory and Techniques, Institute of Radio Engineers, New York, N. Y., vol. MTT-3, no. 2, 1955, pp. 127–33.

3. Coupled Strip Transmission Lines with Rectangular Inner Conductors, **J. D. Horgan.** *Ibid.*, vol. MTT-5, no. 2, 1957, pp. 92–99.

4. The Magnetic Circuit of Telephone Relays, **S. Ekelof.** *Transactions*, Chalmers University of Technology, Gothenburg, Sweden, no. 134, 1953, pp. 1–32.

5. Calculation of the Leakage Flux of the Pendent-Armature Relay, **J. W. Cohen, H. Grosser.** *Philips' Communication News*, Eindhoven, Netherlands, vol. 4, 1952, pp. 125–29.

6. Tables of Integrals and Other Mathematical Data (book), **H. B. Dwight.** The MacMillan Company, New York, N. Y., 1947.

7. Electrosensitive Recording Paper for Facsimile Telegraph Apparatus and Graphic Chart Instruments, **G. Hotchkiss.** *Western Union Technical Review*, New York, N. Y., vol. 2, no. 4, 1948, pp. 176–87.

LATE DISCUSSION

The following discussion was received too late to be included with the paper.

The Logarithmic Diode Counting-Rate Meter and Period Meter

Discussion of paper 57-403 by Bruce B. Barrow, published in Communication and Electronics, September, 1957, pages 421–27.

R. F. Shea (General Electric Company, Schenectady, N. Y.): Mr. Barrow's use of the decilog for growth factor emphasizes the lack of standardization in this field. It must be quite confusing to the neophyte to read of periods in seconds, growth factor in decilogs, doubling time, and decades per minute. I should like to enter a plea here for a concerted effort to adopt one best form, if there be one, and then utilize it. This is not intended as criticism of Mr. Barrow's use of decilog, which I consider quite logical.

In addition to the effects of the various circuit time constants on signal-to-noise ratio and on speed of response, analyzed in the paper, there is a third, very important, function in reactor instrumentation and control: providing a scram or drive-down signal for excessively fast transients (or short periods). The period circuit must be so designed that existence of periods shorter than a prescribed figure will produce this signal if they persist long enough. The duration required must decrease as the period decreases. For example, a drive-down might be required one second after imposition of a 3-second period, but only 100 milliseconds after a 100-millisecond period is initiated. If the integrating time constant is comparable to the differentiating time constant this response time will be lengthened comparatively for fast periods. This consideration usually restricts the integrating time constant to about one tenth of the differentiating time constant. If the response of the period circuit is too sluggish at fast periods the reactor may exceed the danger level during a too rapid start-up.

A similar analysis of the logarithmic-diode smoothing circuit was made in 1953 by H. Hurwitz and E. Clancy of the Knolls Atomic Power Laboratory and reported in internal progress reports. The performance was checked experimentally. The equipment is designed to minimize period overshoot by maximizing K; see equation 27. This is done by making i_0 nearly equal to the minimum detectable current and by minimizing C by the application of negative feedback.

It is interesting to compare Mr. Barrow's analysis with that of Cooke-Yarborough in his 1951 paper,[1] in which the latter describes his system of approximating a wide-range logarithmic characteristic by superposition of a number of saturated characteristics. He also develops an expression for the deviation of the output signal and gives a table showing 21% fluctuation at one pulse per second, 15% at 100 and 4.5% at 10,000. In practice, however, he claims that the standard deviation did not exceed 5% over the 4-decade range. A cursory comparison of these two circuits indicates that both possess the characteristic of incremental resistance inversely proportional to counting rate. From this and other considerations one might conclude that both circuits should have identical transient behavior, if designed to cover the same range with the same response speed. It would be interesting to learn if Mr. Barrow has any information to prove or disprove this.

Finally, in the section "Design of a Low-Level Period Meter," the author conveys the impression that it is impractical to count above 100,000 counts per second, and that resolving times much less than a few microseconds are rare. Nuclear instrumentation has been built which can be used up to 10^6 counts per second, and transistorized equipment is currently under development covering the range of 1 to 10^6 counts per second, with counting loss not exceeding 40%. While, admittedly, circuit design becomes much more difficult, it should be possible to extend this upper limit to possibly ten times this figure. In private communications Cooke-Yarborough has mentioned transistorized scalers counting up to 15 megacycles.

Reference

1. An Accurate Logarithmic Counting-Rate Meter Covering a Wide Range, **E. H. Cooke-Yarborough.** *Proceedings*, Institution of Electrical Engineers, London, England, April 1951, p. 98.

Bruce B. Barrow: In his second paragraph Mr. Shea draws attention to what appears to be an error of omission, the failure to discuss the log-diode period meter as a device to initiate scram upon the appearance of excessively fast transients. It is true that first emphasis in my paper is given to the meter as an instrument for aiding in the controlled start-up of reactors. I do not agree, however, that a meter designed according to the principles set forth in the paper, achieving a balance (suitable for a control application) between high noise and rapid transient response, will be inferior as a detector of dangerous transients to a meter with unequal differentiating and filtering time constants. In Fig. 4 are compared different sets of time constants, all of which produce roughly equal performance as noise filters. The combination suggested by Mr. Shea, with one time constant ten times as great as the other (it makes no difference which is the greater), lies somewhere between case II and case III. Except for the first 10% of the transient, case I, with equal time constants, shows faster response. For the first 10% the difference is too small to be shown on the graph.

It is unfortunate that the Hurwitz-Clancy work was not published so that their contribution might be assessed. I infer from Mr. Shea's comments that these authors succeeded in finding the solution of equation 20 for exponential inputs. It appears, however, that they were working on a different problem, presumably at higher flux levels, since in the log-diode counting-rate meter one neither maximizes K nor minimizes C, which is ordinarily an external capacitance of the order of a microfarad.

The standard deviation in counting rate measured by Cooke-Yarborough on his instrument was found to be about 5% at a counting rate of about 90 pulses per second, and he does not report on measurements taken over the 4-decade range of the instrument. His theoretical results on random fluctuations were applied to a meter with shorter time constants than those of his experimental meter, but in any case the random fluctuations in the Cooke-Yarborough meter should decrease with increasing counting rate. In this respect his meter is more sluggish at high counting rates than the log-diode counting-rate meter, and I should hesitate to conclude that both circuits would have identical transient behavior.

Willamette Valley Radio System for the United States Corps of Engineers

C. PEDERSEN
NONMEMBER AIEE

D. J. MARIHART
ASSOCIATE MEMBER AIEE

WINTER FLOODS in the Willamette River Basin in Western Oregon wreak havoc and destruction from frequent flooding of a highly populated and developed valley, 150 miles in length. Congress instructed the United States Corps of Engineers (USCE) to study the problem in the Willamette Valley and to make specific recommendations. As a result the Corps has recommended a system of 16 multiple-purpose reservoirs. These reservoirs were to be located on the major tributaries in the basin and would control a total of 3,200 square miles of the 11,200 square miles in the basin. At the present time five of the recommended reservoirs are in operation, two are under construction, and planning is under way on two more.

The reservoirs were authorized as multipurpose projects with flood control their primary function. Space is reserved in the reservoirs for flood regulation during the rainy season. Following the flood season the same space is filled and provides stored water for irrigation, navigation, power, pollution abatement, and other benefits. Operation of the reservoirs is directed from the District Office; therefore it is essential that reliable communication circuits be maintained at all times between the District Office in Portland and the dams to insure co-ordinated regulation of all reservoirs. This is particularly important during flood periods. Unfortunately, it is during these periods when the usual communication facilities are apt to be out of service, as most of the dams are serviced by facilities which utilize overhead wires for at least part of their communication circuit. Overhead wires are particularly vulnerable during periods of intense storms. This concern lead to the initiation in 1954 of a radio communication system for the Willamette Valley. By providing an alternate means of contacting the dams there would be less possibility of a complete communications breakdown during a crucial period. Ineffective co-ordination of a system of flood-control reservoirs, which could occur if communications failed, could result in failure to provide adequate control, or possibily even in aggravated damages. Elimination of such a possibility clearly justified the construction of an alternate means of contacting the dams during emergency situations. Commercial telephone service between the District Office in Portland and the dams was retained as an important component part in the communication facilities between the District Office and the flood-control dams.

General System Plan

Original planning for a radio communication system linking the District Office in Portland with the existing flood-control dams, considered the use of low-frequency radio equipment. This type of equipment does not require line-of-sight transmission, and there is no need for repeater stations. However, it does present other problems that could reduce the reliability of such a system. These unfavorable characteristics are skip, fading and a high noise level.

After serious consulation with experienced radio engineers it was decided to install very high-frequency radio equipment. This immediately raised the question of repeater stations, which are expensive to construct and maintain. In the process of investigating these aspects of the radio communication system, the idea of utilizing microwave for the backbone of this radio system was discussed with Bonneville Power Administration (BPA). It was agreed that one audio channel of their microwave system between Portland and Coburg would be available to the Corps of Engineers to facilitate a wireless communication system between the District Office in Portland and the existing flood-control dams in the Willamette Valley. This eliminated the need for constructing repeater stations.

Four of the existing five dams in Willamette Valley are within 35 miles of

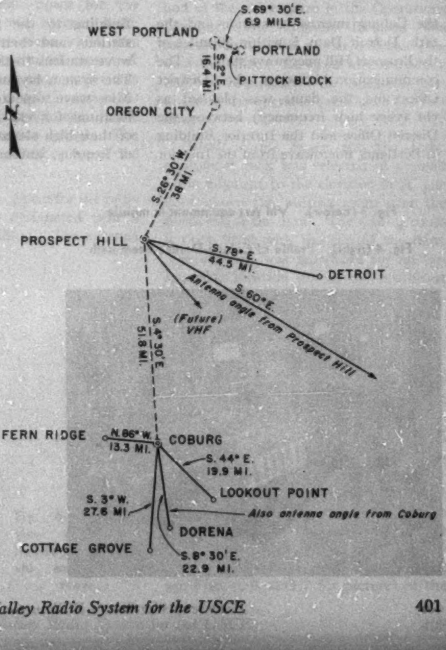


Fig. 1. Station sites and transmission paths: Willamette River Basin, Oregon

—— Vhf path
- - - Microwave path

Paper 57-914, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 20, 1957; made available for printing January 15, 1958.

C. PEDERSEN is with the United States Corps of Engineers, Portland, Oreg.; D. J. MARIHART is with Bonneville Power Administration, Portland, Oreg.

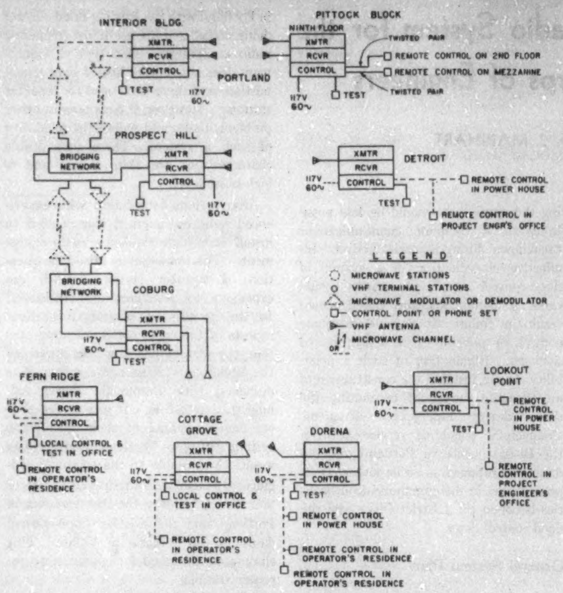

INTERIOR BLDG.
XMTR.
RCVR
CONTROL
TEST
117V 60~
PORTLAND
PITTOCK BLOCK
NINTH FLOOR
XMTR
RCVR
CONTROL
TWISTED PAIR
REMOTE CONTROL ON 2ND FLOOR
REMOTE CONTROL ON MEZZANINE
TWISTED PAIR
117V 60~
TEST
PROSPECT HILL
BRIDGING NETWORK
XMTR
RCVR
CONTROL
TEST
117V 60~
DETROIT
XMTR
RCVR
CONTROL
117V 60~
TEST
REMOTE CONTROL IN POWER HOUSE
REMOTE CONTROL IN PROJECT ENGR'S OFFICE
LEGEND
MICROWAVE STATIONS
VHF TERMINAL STATIONS
MICROWAVE MODULATOR OR DEMODULATOR
CONTROL POINT OR PHONE SET
VHF ANTENNA
MICROWAVE CHANNEL
OR
BRIDGING NETWORK
COBURG
XMTR
RCVR
CONTROL
117V 60~
TEST
FERN RIDGE
XMTR
RCVR
CONTROL
117V 60~
LOCAL CONTROL & TEST IN OFFICE
REMOTE CONTROL IN OPERATOR'S RESIDENCE
COTTAGE GROVE
XMTR
RCVR
CONTROL
117V 60~
LOCAL CONTROL & TEST IN OFFICE
REMOTE CONTROL IN OPERATOR'S RESIDENCE
DORENA
XMTR
RCVR
CONTROL
117V 60~
TEST
REMOTE CONTROL IN POWER HOUSE
REMOTE CONTROL IN OPERATOR'S RESIDENCE
REMOTE CONTROL IN OPERATOR'S RESIDENCE
LOOKOUT POINT
XMTR
RCVR
CONTROL
117V 60~
TEST
REMOTE CONTROL IN POWER HOUSE
REMOTE CONTROL IN PROJECT ENGINEER'S OFFICE

Fig. 2. Block diagram of radio system: Willamette River Basin, Oregon

the Coburg microwave station and the fifth, Detroit Dam, is within 30 miles of the Prospect Hill microwave station. The communication between the District Office and the dams was planned as vhf (very high frequency) between the District Office and the Interior Building in Portland, microwave from the Interior Building to the respective microwave stations, and then vhf from the microwave stations to the flood control dams. The system layout is shown in Fig. 1. Microwave repeater stations are ideal locations for vhf land stations because of their high elevation and the existence of housing, antenna tower, and power facilities. The basic system plan resolved itself as follows:

1. A very low power (5-watt) vhf point-to-point radio link between the USCE District Office and the Interior Building in Portland where the BPA terminates all of its main microwave systems, including the Portland–Alvey system.

2. Full duplex, 50-watt vhf land stations at Prospect Hill and Coburg Microwave repeater stations.

3. The two vhf land stations would be linked to the flood-control dams by means of half duplex, push-to-talk, 50-watt vhf fixed stations at each of the dams. The vhf station at Prospect Hill would serve Detroit Dam and the Coburg vhf station would serve Fern Ridge, Cottage Grove, Dorena, and Lookout Point Dam.

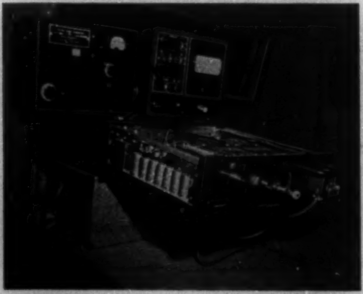
A block diagram of the system as proposed is shown in Fig. 2.

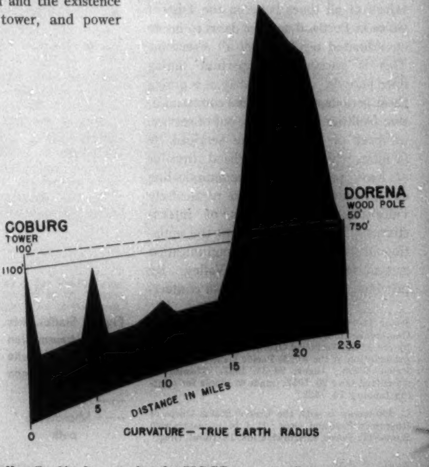
Field Surveys

Vhf field strengths between the USCE dams and repeater stations were established using a jeep with a 25 watt vhf mobile set connected to a 36-foot portable mast and quarter wave whip antenna at the Microwave repeater stations. In most instances the topography between the dams and repeater stations prevented direct line-of-sight communication. A panel delivery sedan equipped with a 25-watt vhf mobile set was used at each of the dam sites to check received vhf signal strength from the microwave stations. The vhf receivers were calibrated with a Measurements Corporation Model *76FM* r-f (radio-frequency) signal generator. (See portable vhf test equipment, Fig. 3.)

Fig. 3 (below). Vhf test equipment in mobile

Fig. 4 (right). Profile of Coburg-Dorena beam path





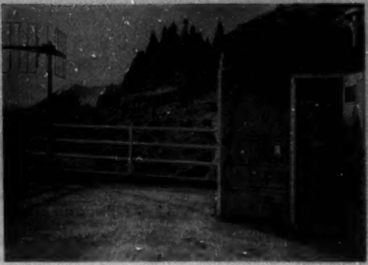


Fig. 5 (left). View of Dorena Dam installation. Gage house door open showing vhf equipment rack and antenna on pipe

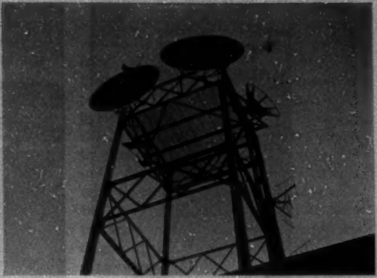


Fig. 7 (right). Closeup view of USCE vhf and BPA microwave parabolic antennas on Interior Building tower

Of the five vhf links to the dams, only the one to Fern Ridge Dam was a line-of-sight path. The other four paths were similar to the Dorena–Coburg path shown on the profile map of Fig. 4. Since the reception at these sites was due to reflected waves, a careful check was made to determine the values of nulls and maximums in the standing wave pattern. The minimum signal strength recorded for any of the Links was 2 microvolts at Cottage Grove. A typical vhf receiver input sensitivity is 0.6 microvolt for 20 db (decibels) quieting. The preliminary surveys proved that the communication network contemplated by the USCE was entirely feasible from an engineering standpoint. In general, it was decided to employ transmitters having a minimum output power of 50 watts and high gain directional antennas at all stations. The antennas at the microwave stations would be semidirectional, having a cardioid radiation pattern, to provide sufficient beam width to cover the geographical spacing of the damsites. It was evident that these factors together with the higher antenna elevations possible in the final installations, would provide considerable improvement over the survey data and would result in an ultimate communication system having a high degree of reliability.

The Willamette Basin communication system was authorized on the basis of the preliminary survey data early in October 1954. Shortly thereafter a joint field survey was made by USCE and BPA engineering personnel to Lookout Point, Dorena, Cottage Grove, Fern Ridge, and Detroit Dams for the purpose of determining equipment mounting locations, power and communication circuit routings. The following plan as shown in Fig. 2 was evolved on the basis of this survey:

Lookout Point Dam. Space for the radio equipment rack was assigned in the small communications room which was directly north of the control room. A 117-volt a-c clock bus power outlet was available there to supply continuous power to the radio equipment even in the event of a station service outage. The antenna was to be mounted on top of the powerhouse.

Cottage Grove Dam. Space for the radio equipment rack was designated near a desk in the office building. The antenna was to be mounted on a 50-foot wood pole 2 feet from the south end of the office building.

Fig. 6. Over-all view of Interior Building showing vhf antennas on tower

Dorena Dam. Space for the radio equipment was reserved in the north end of the reservoir gagehouse on top of the east end of the dam. A combination underground cable and overhead telephone line was planned from the adit house in the dam to the Office Building and resident engineer's house approximately one-half mile distant; see Fig. 5 for a picture of the Dorena Dam installation.

Fern Ridge Dam. Space for the radio equipment rack was assigned in the east end of the office space in the Operations Building. The antenna location was selected on a power service entrance pole near the office. An open wire telephone line was required for an extension telephone in the operator's residence 400 feet away.

Detroit Dam. Space for the radio equipment rack was selected in a small room adjacent to the elevator shaft, one floor above the waiting room near the north end of the dam. The waiting room is on a level with the top of the dam. The antenna was scheduled for an unimposing location above the northwest corner of the elevator shaft structure.

System Design

No special buildings were needed for this system. The system was laid out to use existing buildings in all cases. At the dam sites the equipment was located either in the dam structure itself or in office buildings nearby. At the microwave stations, the vhf repeater equipment was located in the same room with the microwave equipment. The microwave buildings are concrete structures, approximately 27 feet square and 11 feet high; see Figs. 5, 6, and 7 for pictures of the

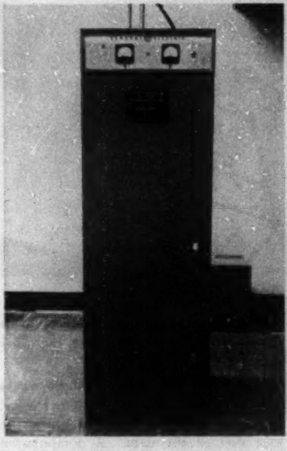


Fig. 8. Front view of vhf base station with cabinet door closed

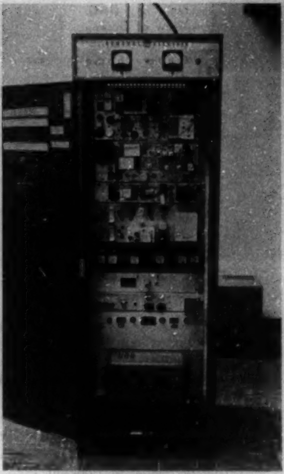


Fig. 9. Front view of vhf base station with cabinet door open

Fig. 10. Vhf base station meter panel

the final installations at Dorena Dam and the Interior Building, respectively.

Only two antenna poles had to be set for the system. In all other cases either existing towers or wall supports were used. The towers at the microwave stations are of the self-supported, lightweight, steel transmission type with top mounted platform. The vhf antennas were located on the sides of these towers.

Equipment Specifications

"Turnkey" type specifications were written requiring the contractor to prepare the necessary detailed design for system components, supply all of the material, install the equipment and energize and test the completed communication system. Fulfillment of the contract required that a satisfactorily operating communication system be turned over to the Corps of Engineers. Land-Air Incorporated of Chicago, Illinois, the successful bidder, supplied the following basic system in accordance with specifications.

Vhf Radio Installation

General Electric 50-watt base station vhf radio equipment was provided for each of the nine locations, including the point-to-point link in Portland between the BPA microwave terminal in the Interior Building and the District Office in the Pittock Block Building; see Figs. 8, 9 and 10 for pictures of the equipment. The vhf base station equipment at the microwave stations and at the Pittock Block Building were equipped with blowers for continuous full duplex operation while the rest of the vhf radio racks at the dam sites were provided without blowers for intermittent, push-to-talk, half duplex operation. It should be noted that since three of the vhf stations, Fern Ridge, Interior Building, and Pittock Block, had line-of-sight paths, the vhf transmitters are operated at a reduced power output of five watts. The six other stations are operated at about 80% peak r-f power output.

Special Features

The vhf system incorporated the following unique features:

1. Base station equipment at the microwave stations and at the Pittock Block Building operate full duplex continuously, using transmit frequency of 163.425 megacycles (mc) and receive frequency of 167.175 mc. The Interior Building vhf station operates on 167.175 mc transmit and 163.425 mc receive. Transmitting and receiving antennas are spaced vertically on a pipe mounting. With this spacing and 4 mc frequency separation, cavities are not required to prevent receiver desensitization.

2. Any station in the system may selectively call any other station by dialing a standard telephone set. This dial, through a relay panel, keys a 3,000 cps (cycles per second) tone generator at the dial rate of 10 impulses per second. The signal is transmitted through the system and is received at each one of the vhf stations on a 3,000 cps tone decoder which operates a stepping switch in accordance with the dial pulses for ringing the called station. The selectors at each station operate every time any call is placed. Only the selected one will operate the telephone ringer, the others will be automatically reset after a 10 second delay. In addition to the selective call at each station, there is a conference call number, 9-0, which will cause all station phones to ring. This, of course, is used for emergency purposes only. The numbers used in this system are relatively simple, utilizing a series stepping arrangement, e.g., 9-1, 9-2, 3, etc. Of course the vhf sets at the dam station, being push-to-talk, have to be turned on during the transmission of the dial pulses. When a connection has been made, any third party may break in at any time in case of emergency. The radio sets at the Pittock Block operate full duplex, hence, there is no push button on the telephone hand set. The bridging network necessary to accomplish the above functions at the Prospect Hill repeater station is shown in Fig. 11. An initial attempt to parallel three hybrids was not successful.

3. A ringback feature is provided at the vhf repeater stations so that a station at the dam may dial the vhf repeater station number; e.g., 9-2, for Coburg and receive an interrupted ringback tone of 60 cycles. Reception of the ringback indicates to the calling party that the vhf link is operating.

4. Another unique feature of this system is that it is designed for a future reception of hydrologic data through telemeter signals of both repeater stations. When these signals come in, they will be in the upper end of the audio range between 2,800 and 3,200 cps.

Components of System Cost

The following percentages, represent the breakdown of costs on this particular job, and give an indication of what may be expected in locations of this nature, for systems of the type outlined:

Item	Per Cent of Total
Field surveys	1.5
Engineering	6.5
Vhf radio equipment (including installation labor cost)	85.5
Test and energization	6.5

In this particular system the cost of land, housing, site development, towers, emergency power, commercial power and roads was negligible since all of these were existing. This, of course, contributed to a very economical over-all system cost.

System Installation and Operation

The tests revealed the following system quality:

Test No. 1. Stability Versus Line Voltage. The Interior Building vhf Station was selected to be spot checked because of the relative ease in varying the voltage ±5% from 117 volts alternating current. The General Electric vhf equipment *r-f* stability was ±0.0003% which is within the specification requirement of ±0.0025%. The audio level stability was better than −2.0 db which is very good.

Test No. 2. Audio Level Settings and Dial Check. The audio level settings based on a 1,000 cps continuous tone were +6.5 dbm at the vhf receiver output. This is 14.5 db above the specifications minimum requirement of −8.0 dbm. All of the stations were dialed at least three times from every other station. Most stations responded on all three calls, some stations responded on two out of three attempts.

Test No. 3. Noise. The average F1A weighted signal to noise ratio of all links tested was 34.9 db with a maximum for one link of 45.8 db and a minimum of 28.0 db. This compared to a specification requirement of 40.0 db; however, it was felt that 40.0 db was too stringent for a vhf system and that a more practical value would be 30 db. The average of 34.9 was considered acceptable.

Test No. 4. Audio Channel Frequency Response. The average deviation between 300 cps and 3,000 cps for all links tested was ±4.4 db which was only 1.4 db above the specification requirement of ±3.0 db. This was considered as a minor exception to the specification and was acceptable.

Test No. 5. Harmonic Distortion. The average harmonic distortion for all links tested was only 5.7% which was well within the specification requirement of 10.0% maximum.

Test No. 6. Vhf Receiver Sensitivity. The received signal strength for each station was greater than the minimum specification requirement of 6 microvolts. Since the maximum receiver sensitivity is 0.6 microvolts for 20 db quieting, a minimum fade margin of 20 db was assured for each station.

The average r-f transmitted power for all stations was 59.6 watts which was well above the specification minimum requirement of 50.0 watts. For normal operation it was recommended that each station be operated at the lowest practical power, e.g., 40 watts, in order to extend tube life and minimize emergency maintenance calls.

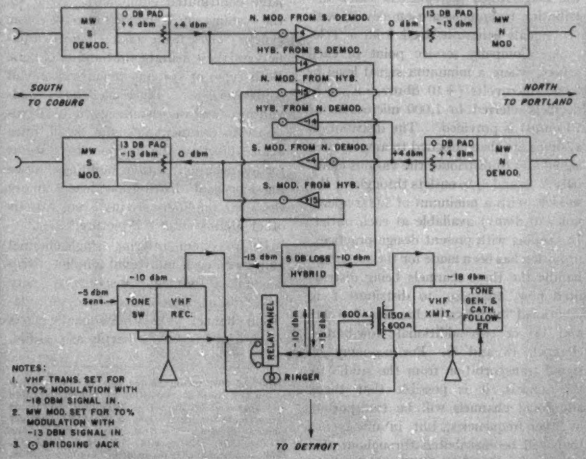


Fig. 11. Block diagram of bridging network at Prospect Hill

Maintenance Logs During Test Period. A total of 18 separate problems were worked on over the two month period from Aug. 27, 1956 to Oct. 27, 1956. Of these 18, only seven were maintenance problems which ranged from weak tubes to blown fuses. The remaining 11 were of the "debugging" type which were solved by permanent modifications being made to the equipment. Considering that nine separate vhf equipment racks comprise the system, only seven maintenance problems over a two month period indicated that good, reliable service can be expected from the system with a normal routine maintenance program.

During the initial phases of putting the equipment into operation and subsequent testing, problems presented themselves as is the case in most systems of this type. These were overcome in the following manner:

1. One of the first problems which had to be solved was the strange behavior of the Coburg to Dorena Dam vhf link. This path was non-line-of-sight as indicated by the beam path profile shown in Fig. 4. The Dorena antenna was first mounted on the north side of the gagehouse on top of the dam structure and oriented for maximum received vhf signal strength. After a month of operation, it was noticed that the signal strength was varying from an unusable minimum to maximum although there was no noticeable movement in the antenna structure. The change in signal strength was finally attributed to the fluctuation in tailwater, which was immediately below the location of the antenna. The tailwater had been reflecting the vhf signal to the antenna, causing extreme fading. The solution to this problem was to move the antenna location ten feet further north and 15 feet higher in elevation where the effect of the reflected signal was minimized.[1,2]

2. The second problem was the one of overmodulation on the microwave channel due to excessive signal level input from the vhf receiver at the repeater station. Trouble was experienced with the station at Prospect Hill because a spurious harmonic from a neigboring vhf station was getting into the USCE vhf receiver and causing the squelch to open. This allowed excessive noise levels of 10 to 12 db above normal at the microwave modulator inputs.

In the original design the microwave input level was to be determined solely by the vhf receiver level output control. This made it relatively easy to overmodulate the microwave since the output control was a continuously variable type of potentiometer. The solution to this problem was to insert 13 db pads on the microwave modulator input circuit. This requires the vhf receiver output level to be 0 dbm and makes it more difficult for an inadvertent high level potentiometer setting on the audio output level of the vhf receiver.

3. The third problem was one of getting the 3,000 cycle decoders to operate properly at the vhf repeater station. A dial repeat relay was added at Coburg to boost the level

of the 3,000 cps dial pulses for proper operation of the dial selectors at Prospect Hill and the Interior Building. Without the dial repeat the tone level was −10 dbm which required the repeater station tone switch sensitivities to be set at maximum and thus make them susceptible to sharp noise transients on the microwave system. The dials at all push-to-talk stations had to be slowed down from 10 impulses per second to 6 impulses per second to permit the Coburg repeat relay to follow the dial pulses accurately. However, trouble was then experienced with the dial repeat relay because of pulse distortion. The final solution was to turn off the selective calling feature at the vhf repeater stations at all times except when maintenance personnel were at these stations. The original 3,000 cps tone from the push-to-talk vhf station is now transmitted directly over the microwave channel to the Interior Building where it is amplified and then transmitted over the vhf link to the Pittock Block.

4. The fourth problem was one of operation wherein the operators who had little or no previous experience with VHF radio equipment tended to talk too loudly into the telephone handsets. This resulted in audio distortion because of clipping by the modulator deviation limiters. Ten db *T* pads were inserted on the vhf transmitter audio input circuit.

5. It was found that the transmitters would drift on power output when operated at their peak efficiency of 60 watts output over an extended period. It was recommended that the vhf transmitter output power be limited to 40 watts at all stations to prevent this trouble from recurring. More efficient blowers may be the permanent solution in a case where it is essential to operate the sets at their maximum power output.

Over-all the system has been working very well for the past year. Two men are used to maintain the entire system of nine stations. These men are located in Portland and have other duties in addition to the Willamette Valley Radio System maintenance. They go out on routine maintenance calls to the stations about once a month. The stability of the tone generators and decoders is very good. Many of the stations have been in operation for over a year with hardly any noticeable frequency drift. This type of system is highly recommended for similar flood control applications.

References

1. VHF PROPAGATION MEASUREMENTS IN THE ROCKY MOUNTAIN REGION, R. S. Kirby, H. T. Dougherty, P. L. McQuate. *Transactions*, Professional Group on Vehicular Communications, vol. PGVC-6, July 1956, p. 13.

2. EXTENDING MOBILE RADIO RANGE BY VHF REPEATERS, C. A. Kemp. *Ibid.*, p. 33.

3. MICROWAVE APPLICATIONS TO BONNEVILLE POWER ADMINISTRATION SYSTEM, R. F. Stevens, T. W. Stringfield. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1088–99.

Television Systems for In-School Teaching

MAX H. KRAUS
ASSOCIATE MEMBER AIEE

THE Teaching-by-Television Experiment now being conducted in Hagerstown, Maryland, under the joint auspices of Electronic Industries Association, the Fund for the Advancement of Education, and the Washington County School District has been described in considerable detail.[1] The early success of this 5-year experiment, which now encompasses 8 schools, with 6,000 pupils, and which eventually will cover all of Washington County, with its 48 schools and 18,000 pupils, is being watched carefully by all those concerned with improving our educational techniques.[2] The plant provided by the Chesapeake and Potomac Bell Telephone Company to link the studio with the schools and the details of the well-equipped studio have been the subjects of separate discussions. This paper will describe the "in-school" or "receiving" portion of the Hagerstown facilities and also will consider possibilities for the future.

System Design

The Television Distribution Systems used in Hagerstown are of the type usually found in schools where signals are received by off-the-air pickup of broadcast stations and distributed from the pickup point to the various classrooms. In the case of the Hagerstown systems, the same distribution techniques are employed, with the signals being obtained from a Telephone Company service point in each school, where a minimum signal level of 3,000 microvolts (+10 dbmv, i.e., +10 decibels referred to 1,000 microvolts at 75 ohms) is provided. The distribution systems are then designed to amplify this signal and to distribute the various channels (2, 4, and 6) to outlets throughout the schools, with a minimum of 1,000 microvolts (0 dbmv) available at each outlet. In keeping with present design practices, provision has been made for the system to handle the three channels being distributed now, but, also, to distribute four additional "high-band" channels (7, 9, 11, and 13) or two additional "low-band" channels (3 and 5). For economy of signal transportation from the studio to the schools, it is possible that these additional channels will be transported at other frequencies; but, in any event, they will be distributed throughout the school using regularly assigned channels so that they can be received on conventional receivers. In this connection, it is recommended that the wiring and other permanent parts of all school television distribution systems be designed to handle all channels (2–13) so that no matter what channels are added, changes will be necessary at only the "head-end."

As might be imagined, several types of distribution systems and master amplifiers can be used, depending on the size of the building in question. In the case of the Washington County Systems, this has been somewhat complicated by the fact that a number of different manufacturers have contributed equipment. Even so, the systems can be divided into two basic types: Those using single-channel amplifiers and those using broad-band amplifiers. Both types of systems are in widespread commercial use. Since each has its advantages and disadvantages, it is appropriate to comment in some detail, once again directing the analysis to systems for distribution within buildings—since for long-haul distribution where many cascaded amplifiers are involved, various other factors enter the picture.[3]

In a system utilizing single-channel amplifiers, each individual amplifier strip has a flat frequency response (usually ±1/2 db) across a 6 mc (megacycle) television channel. On either edge the skirts are designed to fall as sharply as possible.

Paper 57-669, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 16, 1957; made available for printing January 7, 1958.

MAX H. KRAUS is with the Jerrold Electronics Corporation, Philadelphia, Pa.

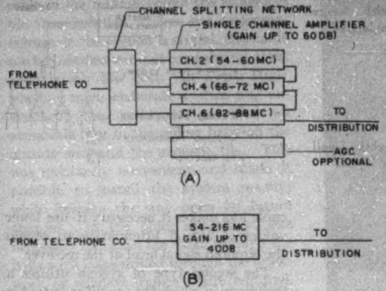


Fig. 1. Amplifier systems

A—Single channel
B—Broad band

The input signal to the master amplifier may be obtained from a common signal source, as is done in Hagerstown (Fig. 1), or each amplifier may be fed from a separate antenna or camera. In either case, the system has the advantage of giving the installer control over the amplification and resultant signal level of each channel, independent of the others. Automatic gain control techniques frequently are applied, and a very flexible system results. Then, too, the relatively narrow amplifier bandwidth makes it possible to obtain higher gains for each stage of amplification than in a broad-band amplifier. Finally, it is possible to obtain both a significantly lower input noise figure and a higher output voltage, since, regarding the latter point, the allowable output voltage in a single-channel amplifier is limited by sync clipping, while in a broad-band multi-channel amplifier, intermodulation between channels gives trouble at a much lower output voltage. Using conventional receiving tubes, it is not unusual to obtain 1.5 volts rms (63 dbmv) per channel from a single-channel amplifier, while limiting the output of a broad-band amplifier to 0.3 volts rms (+50 dbmv). This latter advantage is particularly important in large buildings, such as the new Hagerstown South High School, where a single master amplifier (Fig. 2) is capable of feeding several hundred connection points, even though some eventually may be located over 1,000 feet away from the amplifier.

The broad-band amplifier is at its best in small systems where inidividual channel control and high-output voltage are less important; see Fig. 3. Several types of broad-band amplifiers are available, including chain amplifiers having bandwidths up to several hundred megacycles and others in which two broad-band sections—one covering Channels 2 through 6 (54–88 mc) and another, Channels 7 through 13 (174–216 mc)—are combined in a single unit. In systems of this type (Fig. 1), depending on the type of broad-band circuitry used, the power consumption is reduced, as well as the initial cost and eventual operating costs. Most television distribution system manufacturers have both types of systems available, so that the most desirable one may be chosen for a particular installation.

Before leaving the subject of system design, it should be mentioned that two types of system layouts are possible, with the choice once again determined by the size of the system. Fig. 4 shows a system in which each output of the amplifier is "split" by a resistive (6 db loss) or reactive (3 db loss) splitter to form two trunk lines. From these trunk lines, riser or feeder cables reach into the various areas of the building, with receiver isolation tap-off units inserted at various points to provide lines for the actual set connection. Each set connection point is provided with a matching transformer to im-

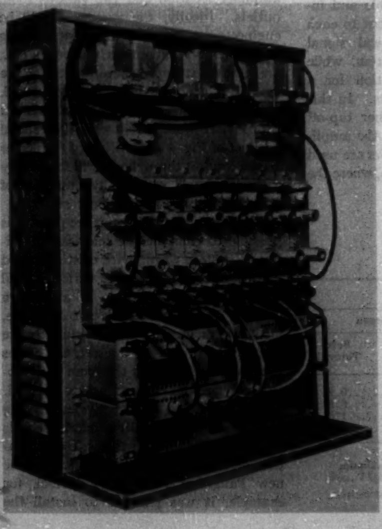


Fig. 2 (left). Typical large master amplifier location

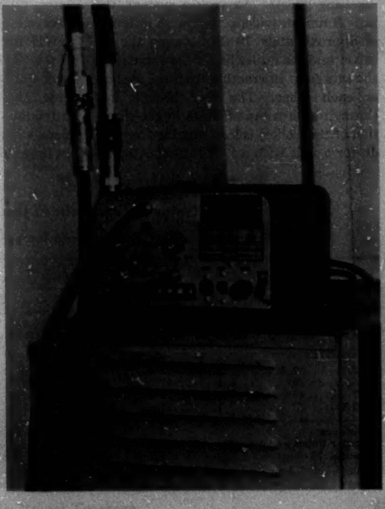


Fig. 3 (right). Small broad-band amplifier location

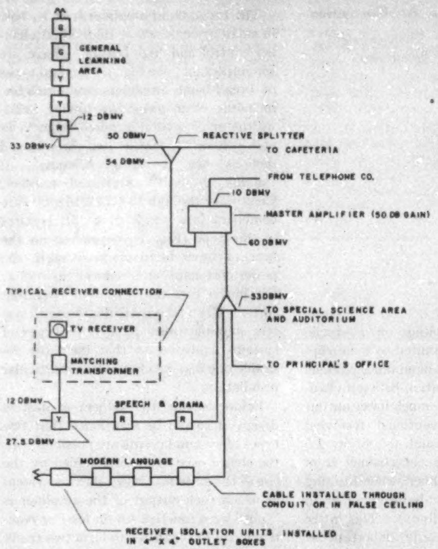
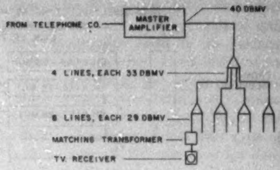

G
G
GENERAL LEARNING AREA
Y
Y
R
12 DBMV
33 DBMV
50 DBMV
REACTIVE SPLITTER
TO CAFETERIA
54 DBMV
FROM TELEPHONE CO.
10 DBMV
MASTER AMPLIFIER (50 DB GAIN)
60 DBMV
TYPICAL RECEIVER CONNECTION
53DBMV
TO SPECIAL SCIENCE AREA AND AUDITORIUM
TV RECEIVER
MATCHING TRANSFORMER
TO PRINCIPAL'S OFFICE
12 DBMV
SPEECH & DRAMA
Y
R
R
27.5 DBMV
MODERN LANGUAGE
CABLE INSTALLED THROUGH CONDUIT OR IN FALSE CEILING
RECEIVER ISOLATION UNITS INSTALLED IN 4" X 4" OUTLET BOXES

Fig. 4 (left). Large system layout

Fig. 5 (right). Small system layout


40 DBMV
FROM TELEPHONE CO.
MASTER AMPLIFIER
4 LINES, EACH 33 DBMV
8 LINES, EACH 29 DBMV
MATCHING TRANSFORMER
TV. RECEIVER

pedance match the 300-ohm receiver to the 75-ohm distribution system. All trunk and feeder lines are terminated by 75-ohm terminating resistors, and every effort is made to provide a matched 75-ohm system throughout. Even the receiver isolation tap-off units are compensated to prevent standing waves and resulting ghosts. In such a system, the tap-off unit provides a minimum isolation of approximately 10 db between the receiver and the feeder line, thus preventing the sets from interacting with the system or each other. The feed-through loss through such a tap-off (loss in the direction of the feeder) is low, ranging from 1.5 db for a unit with a 7 db tap-off attenuation (at Channel 13) to only 0.1 db for a 32.5 unit (at Channel 2). The tap-off units themselves are usually predominately capacitive, so that their tap-off attenuation is less for higher frequency signals than for the lower ones. This compensates for the inherently higher high-frequency loss of the cable and insures a more uniform signal input to each receiver. Fig. 4 shows typical signal levels for a section of a system, while Table I shows the specification for a series of receiver isolation units. In this case, the units with the higher tap-off attenuation are used closer to the amplifiers, while the lower value units are used at the extremities of the system, where the cable loss makes it necessary to use lower values of tap-off attenuation to obtain the required signal level at the reciever.

The second type of system utilizes a technique in which the amplifier output is "split" into four or eight lines, which are then connected directly to the receivers without using receiver tap-off units. This makes it possible to reduce the loss between the amplifier and the ultimate receiver, with a resulting decrease in necessary amplifier voltage; but as shown in Fig. 5, this technique is limited to smaller systems having short cable runs and has the disadvantage that it is difficult to expand the system, while in the former case, additional tap-offs can be added along any feeder cable without disturbing the over-all system in any way.

System Installation

As far as the physical installation of the system is concerned, several approaches are used. In new construction, conduit is generally provided between the chosen amplifier location and the various room outlets. Ideally, the system should be engineered at the time that the building is designed, so that outlets, pull boxes, etc., are located at appropriate points for the maximum economy in time and material. In those cases where this is not possible, a simple riser system, following the general outlines of Fig. 4 and utilizing 3/4-inch or 1-inch conduit throughout, will frequently suffice and will save many hundreds of dollars in future installation costs.

For existing buildings, it frequently is possible to combine the television distribution system wiring with existing sound systems or any one of the various signal services which one finds in a modern school. Where conduit is not available, air shafts or heating ducts frequently are employed or, as a final resort, the cables may be left exposed in the room with protective wire mold used where desired. The Hagerstown School System, being composed of both ultramodern schools and older buildings, utilizes a combination of all of these approaches. In the new Pangborn Elementary School, for example, it was possible to install the

Table I. Typical Receiver Isolation Unit Specifications

Isolation and Feed Through

	Color							
	White		Red		Yellow		Green	
Channel	Isolation	Feed Through	Isolation	Feed Through	Isolation	Feed Through	Isolation	Feed Through
2	32.5	0.1	27.5	0.25	24	0.4	19	0.2
6	29	0.1	27.5	0.25	21	0.4	13.5	0.5
7	25	0.1	22.5	0.25	17.5	0.6	9	1.0
13	23	0.1	21	0.25	15.5	0.6	7	1.5

Impedance
Feeder (input and output);...75 ohms
Tap;...75 ohms

Voltage Sanding Wave Ration
Feeder (input and output);...1.4 maximum
Tap;...1.8 maximum

Bandwidth
Pass Channels 2 through 13

cable in the ceilings before the acoustical tile was installed. Since the maximum voltage in the system is trivial, this cable can be installed without conduit and, thus, at relatively low cost. In South High, where a more elaborate system was envisioned, the architects provided an accessible tray in the ceiling between the outside wall and the acoustic tile. This was extremely convenient and made it possible to install the system readily, while keeping the way open for future changes. In the older schools, it was necessary to install the cable through the halls with holes drilled through the brickwork or window frames where required. Even here, installation can be accomplished speedily using present-day installation techniques.

As can be imagined, the actual direction of the feeder or riser cables is usually in the long dimension of the building. In other words, in multi-story buildings, the rooms are connected from floor to floor, with all rooms in a vertical line being served from the same feeder cable. The modern trend of low, spread-out schools has changed this technique to one in which feeder cable running the length of a hall is used to provide service to the rooms on either side.

Classroom Considerations

Having provided the signals with a path throughout the building, we arrive at the classroom itself. Anyone who has been privileged to watch a class eagerly responding to television teaching is at once struck with the importance of the viewing atmosphere; see Fig. 6. It may be trite to say that the atmosphere must be conducive to learning, but one cannot underestimate the importance of the receiver, its placement, the room lighting, etc.

In a field where hard-and-fast rules have yet to be devised, it still may be of value to others to state some of the "rules of thumb" used in equipping the viewing rooms in Hagerstown. First, it was decided that in the elementary grades, where the students spend most of their time with one teacher in a particular room, all classrooms should be television-equipped so that the students could view their television teaching in familiar surroundings. To provide the desired intimacy between the pupil and the receiver, it was decided to provide sets in the ratio of one set for every 15 students, with a minimum of two sets in each room. According to the standards set by some authorities, this is an unnecessarily generous figure; but since Hagerstown was

Fig. 6 (above). Elementary class watching television

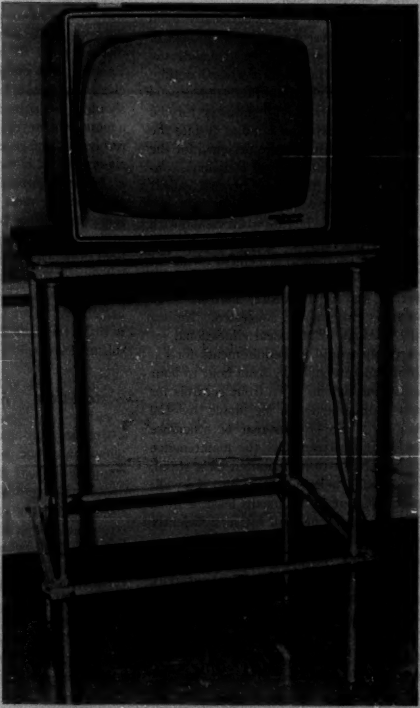


Fig. 7 (right). Television set on stand

designed as a general experiment in learning, it was decided not to attempt to reduce coverage, as might be done in a future test concerned with determining the minimum equipment necessary. In the upper grades, the fact that students are accustomed to moving made it possible in some instances to effect economies by selecting only a few larger television rooms, where four sets have been used. In this latter case, it has been proposed that a study be made to determine the relative learning of students viewing in large groups, as opposed to normal-sized classes. Although this matter must be settled by the educators, rather than by the engineers, it certainly would seem that in terms of flexibility and ability to handle an entire school at times of special-interest programs, the idea of equipping every room has much to commend it. Certainly, even in those cases where, costwise, it is impossible to provide television receivers in every room, the television distribution system should be installed throughout the building, so that reception can be provided easily when required by using moveable sets.

The design of an educationally ideal television reciver could be the subject of an entire paper; but, once again, the experience in Hagerstown is worth noting. A variety of styles and brands of receivers have been provided by industry for the Hagerstown experiment; and despite the fact that almost all were designed for the highly competitive, cost-conscious consumer market, they have given excellent service. Most receivers are 21-inch and give an excellent picture when viewed with normal room illumination if suitable attention is given to placement of the set and to proper blinds or drapes. For the record, and for the future, there are some changes that can be suggested. Among the features of the ideal educational set certainly would be requirements for extreme picture stability from hour to hour and from day to day; front controls for any adjustments to be made by the teacher; tamperproof rear or otherwise hidden adjustments for the maintenance man; ability to be mounted on a stand, on a wall bracket, or built into the wall; and last, but by no means least, a front loud-speaker powered by a sufficiently powerful, high-quality sound system to reach all of the pupils with top-quality sound. In this latter case, a set certainly should have the ability to provide at least 5 watts of undistorted sound.

The television set stand used in Hagerstown was designed by members of the school staff and deserves special mention. As shown in Fig. 7, it combines the features of ease of construction, minimum cost, and portability with extreme ruggedness and functional stability. Three sizes have been used, with heights of 38, 41, and 44 inches. The lower ones have been particularly useful in the lower grades, where, wherever possible, the sets have been placed along the long wall of the room. The lower elevation is particularly desirable when the lesson combines both television viewing and student desk work, since the less amount of head movement between the student's normal writing position and his television viewing position, the better his attention and the less his fatigue when changing from one activity to the other. These particular stands do not provide a downward "tilt" to the receiver, but are easily adapted should such be desired.

As mentioned before, considerable thought has been given to room lighting and to the placement of the television receivers. Ideally, the sets should be placed so that the students can view them without changing from their normal seating pattern and position; furthermore, it is desirable to place the sets along the long wall of a classroom, so that the distance between the television set and the most distant pupil is kept to a minimum. Finally, light coming from outside the classroom should fall over the pupil's left shoulder in the traditional fashion and in a manner to prevent glare on the television receiver. Unfortunately, in many actual classrooms, it is not possible to combine all of the desirable features, so that many compromises must be worked out. In some cases, it has been found desirable to utilize window drapes to block incident light from disturbing the viewing. Room lighting is then supplied by the regular room fixtures to the desired degree. Unlike movie viewing, it is not necessary to make the room light-tight, with the resulting decrease in air circulation, etc.

System Cost

Certainly no analysis of the in-school television system in Hagerstown would be complete without mention of cost. The distribution systems themselves vary in complexity, from those serving a few rooms to the potentially elaborate system in South High School. In other areas of the country, the number of channels, the type of building construction, and local factors will affect the individual costs. In general, however, experience has shown that a complete television distribution system for a school of from 10 to 60 rooms can be provided at a cost ranging from $25 to $50 a room. This figure includes all labor and material for the system, except for the cost of any conduit provided as part of the original building. The cost of the television set, etc., must be added to this system cost. The stand for the television set adds approximately $20 to $25, while a 21-inch receiver can be purchased for between $125 and $150. Based on present labor and material costs, therefore, it is safe to allow between $175 and $200 as the total cost for each installed television-equipped room.

Future Possibilities

Thus, in Hagerstown, we find television systems that are providing low-cost, reliable television viewing to some 6,000 pupils, with planned expansion to over 18,000 students. Already the over-all program has proven itself to have a tremendous potential in improving the very basis of American education. At such a time, it seems appropriate to point out that while the Hagerstown facilities already have attained a high degree of education acceptance, they have actually been in use for only a short time and that there are many technical features which remain to be tried. Many other experiments have been conducted,[4,5] and many groups of statistics have been collected. Although it would be both impossible and pointless for the Hagerstown work to duplicate these other studies, there are some areas which deserve attention. To date, for example, only conventional-size television screens have been used in Hagerstown. Certainly, an experiment with as broad a base as that being carried on in Hagerstown deserves the chance to try large-screen projection as an educational tool. Recently developed equipment would seem to make it possible to utilize such equipment effectively for auditoriums and other large assembly areas.

System-wise, to date, all of the teaching in Hagerstown has been done from a single location. Certainly there are teaching situations where the reduction in the number of pupils being taught will be desirable and where it will be possible to introduce actual voice communication between students in viewing rooms and the teacher. It is possible that this approach may be more applicable to the college atmosphere than to high school and elementary education; nevertheless, it is an approach that is technically possible and is being explored elsewhere.

In a similar vein, the system now installed in South High School has the ability to provide "Two-Way" television[6,7] so that if, at sometime in the future, it becomes desirable to originate programs from the various classrooms, laboratories, or shops of this high school to be viewed at other locations in the school system, the facilities are already available.

Expanding on this concept, it is possible to think of the coaxial cables as broad streets, over which many types of programs and signals can be carried, so that by introducing tone-control equipment in other portions of the radio-frequency sprectrum, it will be possible to give individual instructors the use of control channels to direct experiments or facilities at other locations, which, by virtue of size or distance, may be best integrated into the teaching program by way of a television link.

Finally, in this atmosphere of technical exploration, one may challenge the basic television standards themselves. Certainly, in a completely closed-circuit system, there is no technical or legal necessity for using conventional broadcast standards for picture transmission. There may be instances where, because of requirements for greater details, it will be desirable to use wide-band systems that provide additional definition. On the other hand, possible economies of equipment can be achieved by using narrow-band systems with high-persistence screens for the display of charts and supplementary material where motion is not required. All of these techniques are now technically possible through the use of equipment already available; and while some are being evaluated by educators, many others remain untouched.

Conclusions

In summary, I should like to express my appreciation to the electronics industry, the Fund for the Advancement of Education, the Chesapeake and Potomac Bell Telephone Company, and the Washington County School System, who have made the Hagerstown experiment possible and who have assisted in preparing this paper, and, at the same time, to urge that additional groups of educators and engineers continue to work together to further solidify this marriage of electronics and education.

References

1. **Teaching in Washington County, 1956–1957.** Board of Education, Hagerstown, Md., Sept. 4, 1956, pp. 1–62.

2. **Schools for Tomorrow: An Educator's Blueprint, Alexander J. Stoddard.** The Fund for the Advancement of Education, New York, N. Y., 1957, pp. 9–60.

3. **Line Amplifiers for Community Television Systems, K. A. Simons, Don Kirk, H. J. Arbeiter.** *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 314–20.

4. **An Inventory of Instructional Television Research, Hideya Kumata.** Educational Television and Radio Center, Michigan State University, Ann Arbor, Mich., Dec., 1956, pp. 1–115.

5. **Educational Television Research Fact Sheet Service.** National Association of Educational Broadcasters, Urbana, Ill. (continuing series)

6. **Developments in Closed-Circuit Television, M. H. Kraus.** *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 704–09.

7. **Two-Way Closed-Circuit Educational Television, John R. Martin.** *Research Report no. 948-1 & 2*, Case Institute of Technology, Cleveland, Ohio.

Transient Response of Impulse Voltage Dividers

F. A. FISHER
ASSOCIATE MEMBER AIEE

WITH modern test instruments it is possible to measure the transient response of an impulse-voltage-divider system directly and with good accuracy. The response to a step function, a wave which rises to its steady-state amplitude instantaneously, is particularly easy to determine. This is fortunate since a good approach to understanding the transient response of a complicated system is through study of the step response. The purpose of this paper is threefold: to explain the significance of the step response, to relate the step response to the over-all transient response of the system, and to relate the crest accuracy and step response of a divider system by means of simple accuracy curves. Certain methods which may sometimes be used to improve the response of practical impulse voltage dividers will also be discussed.

Sources of Error in Impulse Voltage Dividers

Most impulse voltage dividers presently in use fall into one of four categories: resistance dividers, shielded or unshielded, capacitance dividers, and mixed dividers. Although this paper is primarily concerned with resistance dividers, much of the material applies equally well to capacitance dividers.

Fig. 1, a schematic diagram of a common resistance-divider system, shows the elements that affect the transient response the most. While such a diagram is well known and the elements easily recognizable, several tacit assumptions are often made when analyzing such a system. First, attention is frequently focussed on only the high-voltage arm neglecting the effect of connecting leads and the low-voltage arm. Second, all the error is assumed to be caused by uncompensated charging currents through the stray ground capacitance of the high-voltage arm. The effects of inductance and through capacitance are usually neglected since they make it difficult to solve the differential equations describing the high-voltage arm.[1] Third, it is assumed that the low-voltage arm should be a pure ohmic resistor, or perhaps that it should have the same L, R, and C characteristics as the high-voltage arm.

Since the purpose of the divider is to produce a replica at the oscilloscope of the voltage being measured, and since each of the elements may cause errors, the

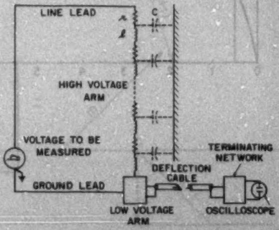


Fig. 1. Resistance voltage divider system

Paper 58-37, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 14, 1957; made available for printing November 4, 1957.

F. A. Fisher is with the General Electric Company, Pittsfield, Mass.

The author wishes to acknowledge the guidance and valuable suggestions of J. H. Hagenguth, A. F. Rohlfs, and J. S. Kresge.

term "divider" in this paper will refer to the entire measuring system including the connecting leads. The primary attention will be given to the performance of the over-all system.

Relationships Between the Step Response and the General Transient Response

The over-all response of a system may be specified in several ways, any one of which is sufficient to describe the complete transient and steady-state response of the system. The response may be specified in terms of the steady-state frequency and phase-shift characteristics,[2,3,4] in terms of the response to an impulse (a wave of infinite amplitude and enclosing a net area of unity), or in terms of the step function response.[5,6] Of the three, the response to a step appears the simplest to relate to the over-all accuracy of the system.

The step response is related to the general transient response by the superposition integral (also called the convolution or Duhamel's integral),

$$h(t)=\int_0^\infty f(\tau)\frac{d}{dt}[\phi(t-\tau)]d\tau \qquad (1)$$

where

$h(t)$ = response to the input
$f(\tau)$ = arbitrary input
$\phi(\tau)$ = response of the system to a step function
$\phi(t-\tau)=\phi(\tau)$, but reversed left to right and shifted to the right by an amount t

An alternative form of equation 1 is

$$h(t)=\int_0^\infty \frac{d}{dt}[f(\tau)]\phi(t-\tau)d\tau \qquad (2)$$

If the step response of the system is known, equation 1, equation 2, or any of the other forms presented in the literature, may be used to find the response to an arbitrary input wave, depending on which form is easiest to evaluate.[7,8]

In general, the step response of an actual high-voltage measuring system cannot be expressed as a simple analytic function and the input wave generally has discontinuities and irregularities which make analytic evaluation of the superposition integral impractical. The integral may be evaluated numerically with the aid of a desk-type calculator however. This is done simply by considering the input wave to be composed of a series of successive step functions, each shifted in time by a constant interval, and then taking the sum of the step response amplitudes at the appropriate time. If the the step response and the applied wave are approximated as straight lines (Simpson's Rule) the bookkeeping may be done in the systematic manner[9] shown in Fig. 2. Figs. 2 and 3 illustrate the principles described. The number of steps into which the applied wave must be broken depends on the waveshapes of the applied and step response waves and on the degree of the detail desired in the calculated response wave. Generally, 30 to 50 steps are sufficient. If only a portion of the response waveshape is desired, only the summation calculations for that portion need be done. The summation calculation for each point is independent of the calculations for other points. A previous paper (reference 5, Fig. 17) gives an example of the accuracy that may be realized by such a numerical method.

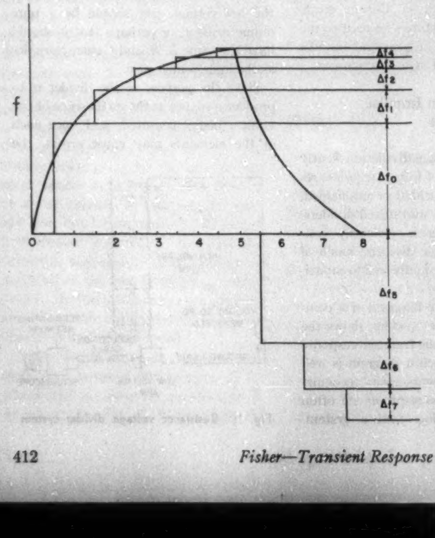
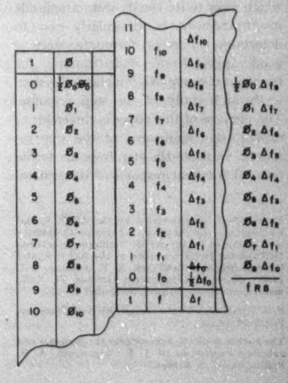


Fig. 2 (left). Representation of a wave by a series of step functions

Fig. 3 (right). Numerical evaluation of a superposition integral

ϕ = amplitude of step response
f = amplitude of applied wave
f_R = amplitude of response wave



Response of Systems to Typical Impulse Waves

Next to a step function the ramp function, a function which increases indefinitely at a constant rate, is the simplest type of applied wave. This is particularly fortunate since many common waves may be visualized as combinations of ramps of various slopes or ramps and steps.

Since the ramp is the integral of the step it can be determined from the step response quite easily by numerical integration. In addition all ramps are related, since changing the slope by a factor *m* merely changes the amplitudes of the points on the response by a factor *m*.

Fig. 4 shows the step responses of four hypothetical systems and their corresponding ramp responses. In each case it will be noted that after the step response has reached its steady-state value the ramp response follows parallel to the applied ramp wave, but displaced by a time T. Hyltén-Cavallius and Romano[4,6] have pointed out that the time displacement T is equal to the net area enclosed between the step and the step response, considering areas below the unity axis positive and those above it negative. In system A, for instance, the time displacement T is equal to the area of the triangle A_1 ($A_1=1/2\times0.1\times1=0.05=T$).

Fig. 5 shows how some common waves may be visualized as combinations of ramps and steps. Since these waves are similar to those encountered during high-voltage impulse testing they will be called

Vol. 7
Pt. 1
1958

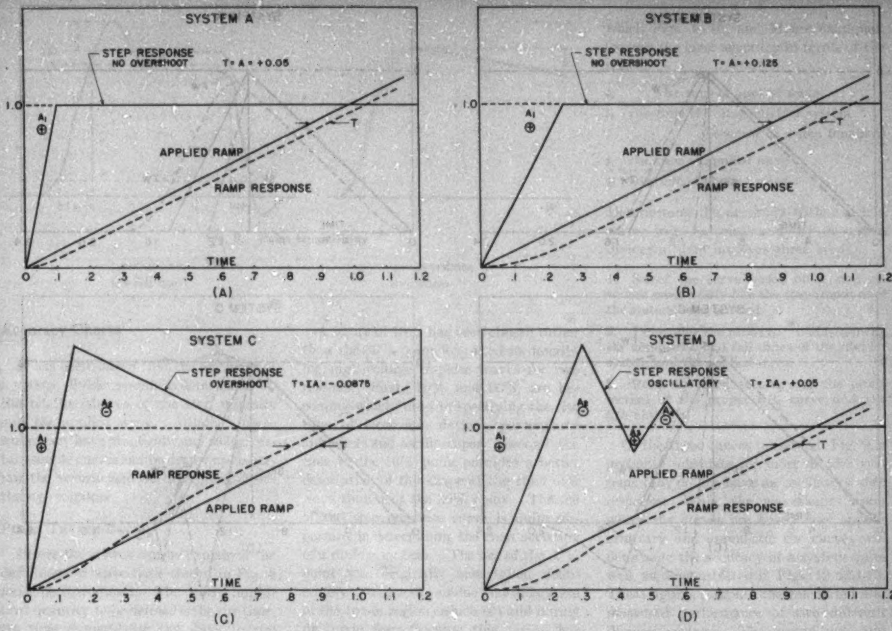


Fig. 4 (above). Step and ramp function response waves

full, triangular, and steep-front waves. Fig. 6 shows how three typical waves would be reproduced by the four systems of Fig. 4. Every time a change in slope or discontinuity is made in the applied wave, a disturbance is set up which lasts as long as the time that would be required for the step response to reach its steady-state value.

Often a divider is rated upon its ability to reproduce a steep-front wave with the tacit assumption that if it can do this correctly it will also reproduce any other wave correctly. Figs. 4, 6, and 7 illustrate that if an understanding of divider response is desired, more than just the crest accuracy of the response wave must be considered. They also illustrate the pitfalls of considering the response to only one type of wave.

Systems *A* and *D* of Fig. 4 both reproduce the steep-front wave with the same crest error, 5%, but the over-all shape of the response wave is quite different; Figs. 6(A) and 6(D). If studies were being made of some phenomena where the shape of the applied voltage wave was quite important, one could arrive at completely different conclusions depending on which divider system was used.

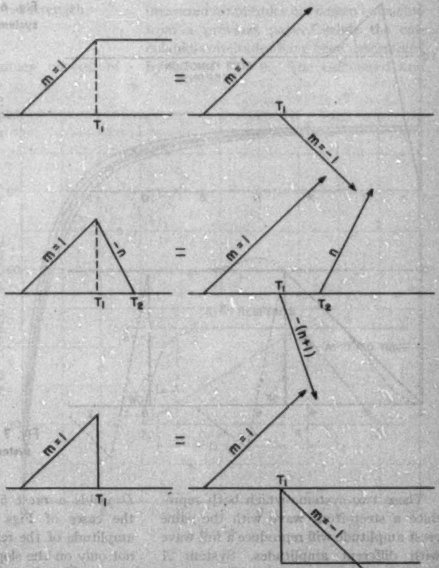


Fig. 5 (right). Representation of waves by ramp and step functions

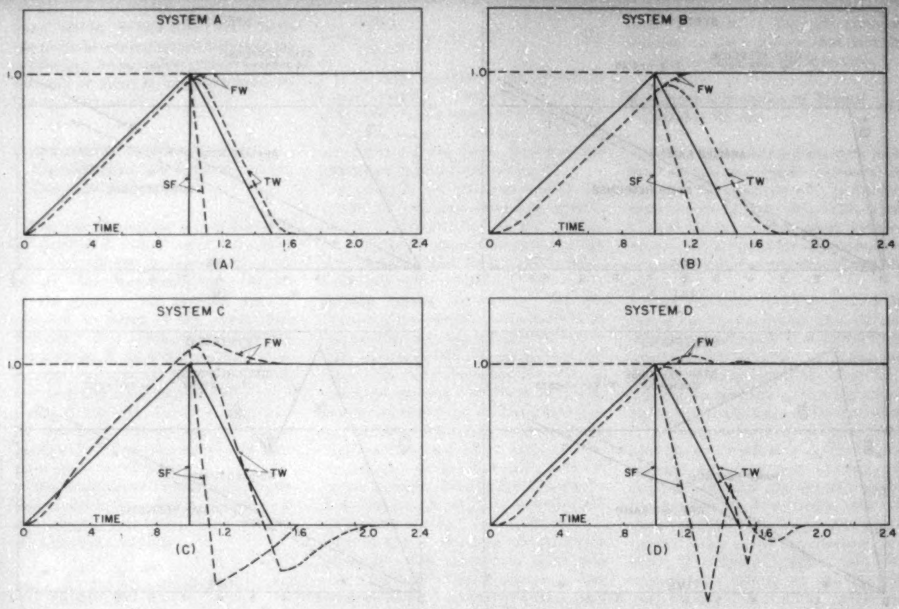


Fig. 6 (above). Response of systems of Fig. 4 to typical waves

——applied wave
– – –response wave
SF = steep-front wave
TW = triangular wave
FW = full wave

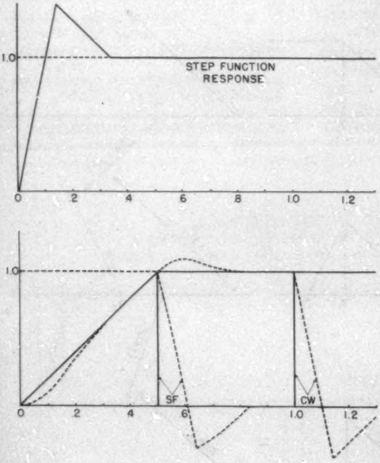


Fig. 7 (left). Response of a system in which ΣA = T = 0

These two systems which both reproduce a steep-front wave with the same crest amplitude will reproduce a full wave with different amplitudes. System *A* yields the correct amplitude, but system *D* yields a crest 5% too high. In all the cases of Figs. 4 and 6 the crest amplitude of the response wave depends not only on the slope of the front of the applied wave, but on the slope of the tail as well. For instance, the triangular wave is always reproduced with a greater crest amplitude than the steep-front wave.

Considering idealized steep-front waves, the crest error at chop time is equal to the vertical displacement between the ramp function and the ramp response of the divider system. In terms of the factors *A* and *T* of Fig. 4, the crest error is

$$\text{error} = 100\frac{T}{t}\%$$

where t is the chop time of the steep-front wave. Some investigators have implied that divider systems should deliberately be made oscillatory so that the total area between the step function and its response be zero ($\Sigma A = T = 0$), thus resulting in a zero crest error when measuring idealized steep-front waves. Fig. 7 illustrates a system with such a response. Although the steep-front wave is measured with no error the full wave or the wave chopped on the tail is measured with a crest 10% too high. If such a system were used to measure volt-time curves the indicated curves would be 10% high in the region where the applied wave was being chopped on the tail.

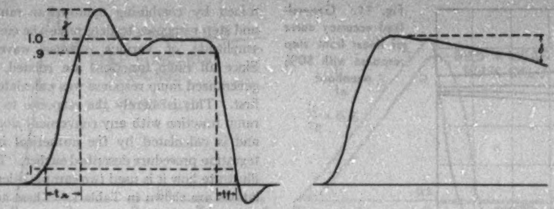


Fig. 8. Pulse terminology

t_r = rise time
t_f = fall time

γ = overshoot
δ = droop

Accuracy Charts

It has been shown that the accuracy of a voltage-divider system is determined by the relative shapes of the step response and the applied wave. Although these waves may have practically any shape, certain simple curves can be drawn up to permit the accuracy to be determined from the step response.

Pulse Terminology

Before the curves can be explained the definitions of waveshape shown in Fig. 8 must be introduced. The most important quantity to be defined is the rise time, the time required for the wave to rise from 10% to 90% of its final amplitude. The fall time is defined similarly. The terms overshoot and droop are self-explanatory.

A figure of 10% has been chosen rather than the 30% point long used in describing high-voltage impulse waves for two reasons. First, 10% and 90% are becoming widely used in specifying the rise time of electronic devices, particularly amplifiers and oscilloscopes. Second, the time to the 10% point provides a better description of the shape of the start of a wave than does the 30% point. The toe of the step response curve is quite important in determining the crest accuracy of a divider system. The use of the 30% point was originally undertaken deliberately to avoid undue consideration of the toe-in region, which is valid during dielectric tests because this region has little influence on dielectric strength.

Use and Significance

The generalized accuracy curves, of which Figs. 9, 10, and 11 are examples, express the crest accuracy in terms of the ratios:

$$\frac{t_1}{t_r} = \frac{\text{rise time of applied wave}}{\text{rise time of voltage-divider system}}$$
(response to a step function)

$$\frac{t_1}{t_2} = \frac{\text{rise time of applied wave}}{\text{fall time of applied wave}}$$

Determining the accuracy with which a wave will be measured by a given divider system involves three steps.

1. Select the curve based on a step response most nearly like the step response of the system being studied.

2. Determine the ratios t_1/t_r and t_1/t_2 using the actual rise and fall times of the divider system and the applied wave.

3. Read the crest accuracy at the intersection of the proper t_1/t_2 curve and the t_1/t_r abscissa.

Of the three curves presented, Fig. 9 is probably adequate for most divider systems that do not have an oscillatory step response. While the waveshapes upon which the curves are based may appear arbitrary and unrealistic the curves will determine the accuracy of a system quite well, as demonstrated in Figs. 12 and 13. These figures compare the calculated and measured performance of two different divider systems. The waveforms and measured amplitudes have been extracted from a previous paper,[8] while the calculated amplitudes have been determined by use of Fig. 9. The calculated and

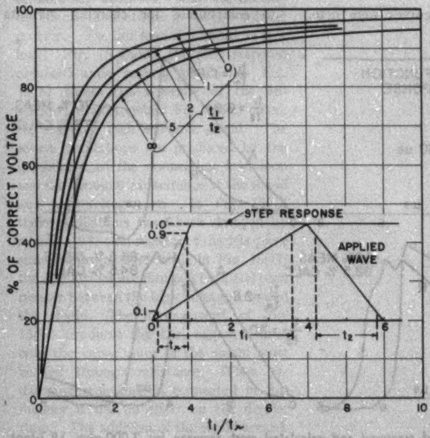


Fig. 9. Generalized accuracy curve

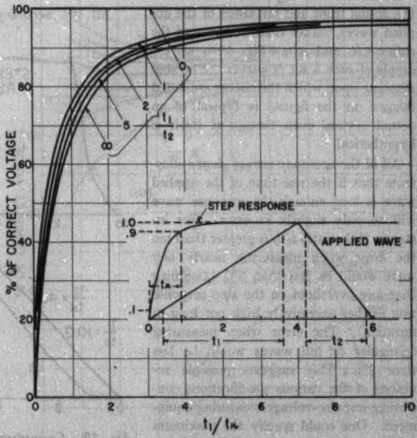


Fig. 10. Generalized accuracy curve of exponential front step response

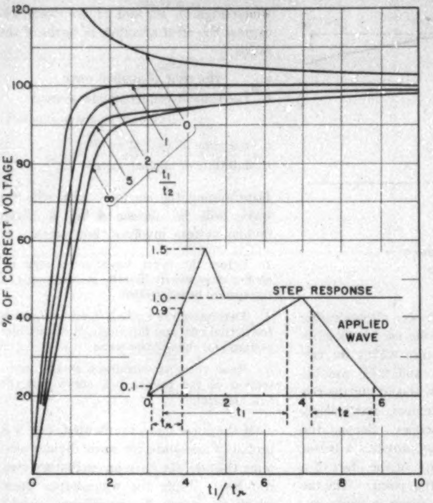


Fig. 11. Generalized accuracy curve of linear front step response with 50% overshoot

measured crest amplitudes agree quite closely in all cases.

Figs. 9, 10, and 11 are suitable for general use since they are presented in terms of the dimensionless ratios t_1/t_r and t_1/t_2. If the response of an actual system has been determined curves may be drawn up which are more accurate, but are valid only for that particular system. Such curves could be presented in terms of the actual front and fall times of the applied waves, rather than in terms of the ratios t_1/t_r and t_1/t_2. Fig. 14 is an example of such a set of curves. The step response upon which the curves are based (shown on the figure) is typical of an actual system even though it itself is hypothetical.

All of the accuracy curves shown illustrate that if the rise time of the applied wave is long enough the response wave will be only slightly in error. For instance, if the ratio t_1/t_r is greater than ten the error when measuring nearly any wave would be less than 5% (assuming that any overshoot on the step response was neither excessively high nor long in duration). The error when measuring triangular or full waves would be less than 3%. This suggests possible revisions of the various specifications concerning impulse-voltage-measuring equipment. One could specify the maximum allowable rise time and overshoot that a measuring system would be allowed for use in measuring a given set of impulse waves or one could specify the minimum rise and fall times of an applied wave that could be measured with a piece of equipment having a given rise time. Either set of specifications would assure that the amplitude of the indicated response wave would be within the allowable limits of error.

CALCULATION OF CHARTS

All the accuracy curves were determined by combining appropriate ramp and step responses to determine the crest amplitude of various reponse waves. Since all ramp functions are related, a generalized ramp response was calculated first. This is merely the response to a ramp function with any convenient slope and is calculated by the numerical integration procedure described earlier. To illustrate how it is used two sample calculations are shown in Table I. These are calculations for the two points, $t_1/t_r=5$, $t_1/t_2=\infty$ and $t_1/t_r=5$, $t_1/t_2=1$, of Fig. 9. The amplitudes of the step function response and the generalized ramp response are those of Fig. 4(A). Calculating the response amplitude at other points is just a matter of combining ramp functions of the proper slope, as shown in Fig. 5.

RESPONSE OF CERTAIN PRACTICAL DIVIDERS

While it has been shown[5] that uncompensated resistance dividers are adequate for measuring the types of impulse waves encountered during acceptance tests on most apparatus it is sometimes necessary to build divider systems with higher performance capabilities. The following section will show how the performance of various compensating devices may be evaluated in terms of their effect on the step function response. All of the step-function-response oscillograms shown were obtained by the methods outlined in the Appendix.

The most common type of compensating device is a shield near the line end to compensate for charging currents

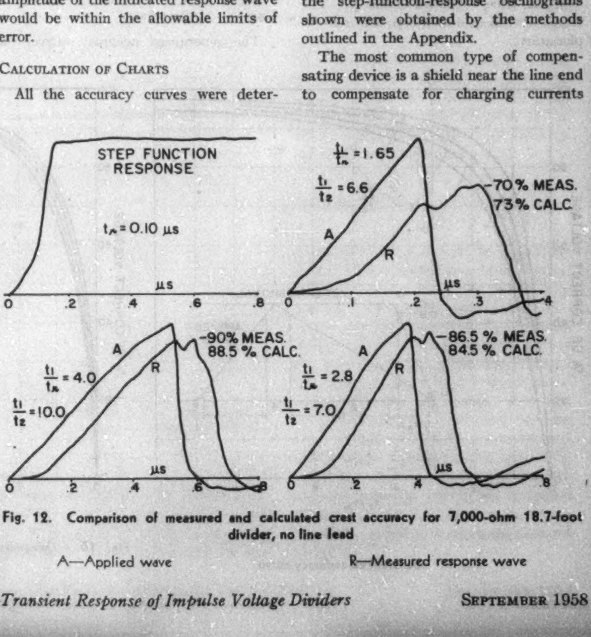


Fig. 12. Comparison of measured and calculated crest accuracy for 7,000-ohm 18.7-foot divider, no line lead

A—Applied wave R—Measured response wave

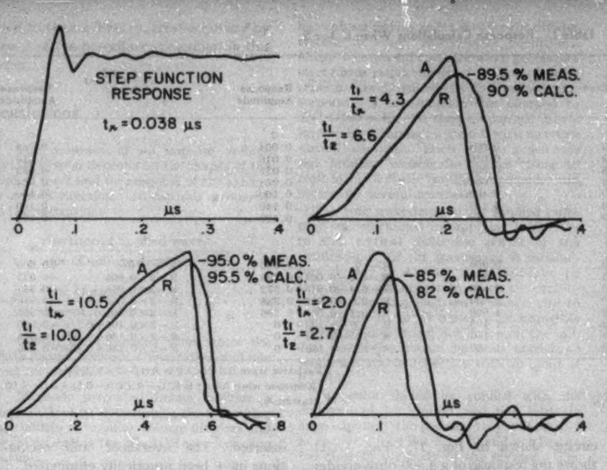


Fig. 13. Comparison of measured and calculated crest accuracy for 7,000-ohm 18.7-foot divider, 9-foot line lead

A—Applied wave R—Measured response wave

through the ground capacitance. To show the effects of a line shield on the step response, a 21,000-ohm divider was assembled. The high-voltage arm consisted of six vitreous enameled resistors $8^1/_2$ inches long by 3/4 inch in diameter with essentially noninductive windings. The low-voltage arm was simply the 75-ohm deflection cable and its noninductive terminating resistors, thus eliminating the effect of inductance in the low-voltage resistor. The line lead was about 14 feet long. Fig. 15 shows the step response obtained under several conditions. Fig. 15(A) shows the response of the unshielded divider, while Fig. 15(B) shows the response with a simple shield. The severe oscillations are produced by the inductance of the connecting lead loop and the terminal capacitance of the shield and high-voltage arm. In Fig. 15(C) these oscillations have been damped by adding a 500-ohm resistor at the test-piece end of the line lead while in Fig. 15(D) the oscillations are damped by a 500-ohm resistor between the high-voltage arm and the shield. The arrangement of Fig. 15(D) produces the best response since only the capacitance of the shield must be fed through resistance. When the damping resistance is connected in this manner it has no effect on the divider ratio. The addition of the shield according to the arrangement of Fig. 15(D) improved the rise time from 0.091 μsec (microsecond) to 0.023 μsec. If a restriction were made that a steep-front wave must be measured with no more than 5% error the unshielded divider would be suitable for waves cresting at 0.91 μsec (t_1=0.73 μsec) while the shielded divider of Fig. 15(D) would be suitable for waves cresting at 0.28 μsec (t_1=0.23 μsec). The performance of the unshielded divider is best estimated from Fig. 10 and the performance of the shielded divider of Fig. 15(D) is best estimated from Fig. 9.

The response of a divider may sometimes be improved by the use of suitable wave-shaping networks in the low-voltage arm. If the response voltage, as developed across a pure ohmic resistor, rises exponentially as in Fig. 16(A) the step response may be improved by inserting an inductor in series with the low-voltage resistor in much the same way as the response of a video amplifier may be improved by the use of peaking coils. Figs. 16(A) and 16(B) show the response of a 10,000-ohm divider suitable for 3,500 kv. The high-voltage arm consisted of 67 resistor cards each holding 8 inches of resistance winding. The total inductance of the high-voltage arm was about 270 microhenrys and the total capacitance to ground about 150 micromicrofarads. Since the divider system was to be used with a short line (15 feet) lead the high-voltage arm was folded near the center and the top portion run horizontally until it was close to the pulse generator.

Fig. 16(A) shows the response with the output of the high-voltage arm fed directly into the deflection cable. Since the output current of the high-voltage arm is essentially independent of the resistance of the low-voltage arm the same response would be developed if a pure ohmic resistor were used as the low-volt-

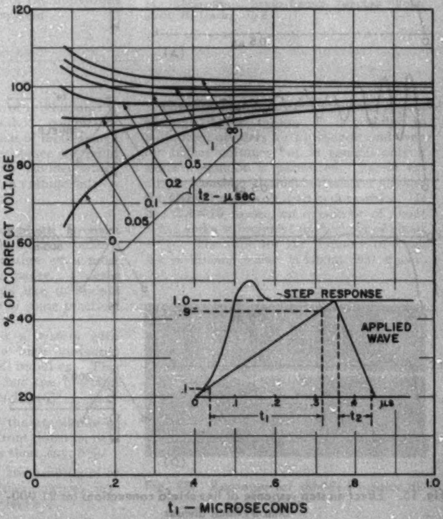


Fig. 14. Accuracy curve for a specific system

Table I. Response Calculations When $t_1/t_r=5$

Time	Step Response ϕ	Generalized R*	R†	‡ $\frac{t_1}{t_2}=\frac{0.5}{0}=\infty$	Response Amplitude	§ $\frac{t_1}{t_2}=\frac{0.5}{0.5}=1$	Response Amplitude
0	0	0	0		0		0
0.02	0.2	0.002	0.004		0.004		0.004
0.04	0.4	0.008	0.016		0.016		0.016
0.06	0.6	0.018	0.036		0.032		0.032
0.08	0.8	0.032	0.064		0.064		0.064
0.10	1.0	0.050	0.100		0.100		0.100
0.12	1.0	0.070	0.140		0.140		0.140
0.14	1.0	0.090	0.180		0.180		0.180
⋮	⋮	⋮	⋮	⋮		⋮	
0.50	1.0	0.450	0.900	R— 0	=0.900¶	R— 0	=0.900
0.52	1.0	0.470	0.940	R−0.2−0.004	=0.736	R−2×0.004	=0.932
0.54	1.0	0.490	0.980	R−0.4−0.016	=0.572	R−2×0.016	=0.948
0.56	1.0	0.510	1.020	R−0.6−0.032	=0.388	R−2×0.036	=0.948¶
0.58	1.0	0.530	1.060	R−0.8−0.064	=0.196	R−2×0.064	=0.948
0.60	1.0	0.550	1.100	R−1.0−0.100=	0	R−2×0.100+0	=0.900
0.62	1.0	0.570	1.140	R−1.0−0.140=	0	R−2×0.140+0.004	=0.864
0.64	1.0	0.590	1.180	R−1.0−0.180=	0	R−2×0.180+0.016	=0.836

* Response to a ramp reaching unity at $t=1.0$.

† Response to a ramp reaching unity at $t=0.5$.

‡ Response when $0.5<t$ is $R(t)-\phi(t-0.5)-R(t-0.5)$.

§ Response when $0.5<t<1.0$ is $R(t)-2\times R(t-0.5)$.
(Response when $1.0<t$ is $R(t)-2\times R(t-0.5)+R(t-1.0)$.

¶ Maximum.

age arm. Fig. 16(A) shows the response when a 5-ohm noninductively wound resistor in series with an inductance of 0.4 microhenry was used as the low-voltage arm. The rise time was improved from 0.33 µsec to 0.14 µsec by the addition of 0.4 microhenry in the low-voltage arm.

If a divider system is underdamped and has an oscillatory step response the general shape of the response may be improved (but not the rise time) by the circuit shown in Fig. 17. Fig. 17(A) shows the response of a 4,500-ohm divider with a high-voltage arm consisting of resistors each containing 8 inches of resistance winding. The line lead was 14 feet long. Fig. 17(B) shows the response with the wave-shaping network inserted. The overshoot and oscillations have been practically eliminated.

The effects of such modifications or compensating features cannot be calculated practically. Their effect on the step function response may, however, be determined experimentally and thus re-

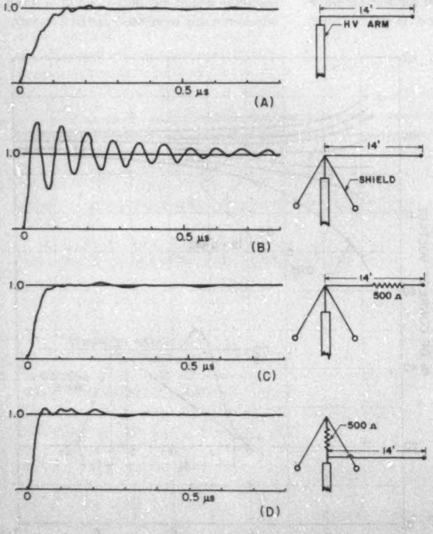


Fig. 15. Effect on step response of line shield connections for 21,000-ohm 51-inch divider

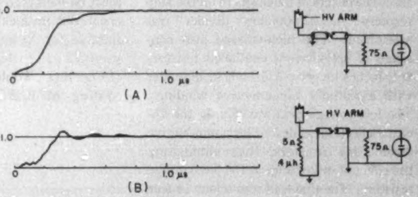


Fig. 16. Effect on step response of inductance in low-voltage arm

A—Noninductive low-voltage arm
B—Inductive low-voltage arm

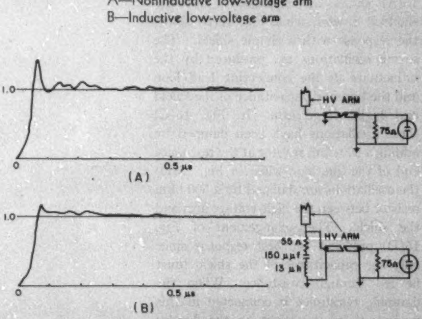


Fig. 17. Effect on step response of a wave-shaping network in low-voltage arm

A—Without wave-shaping network
B—With wave-shaping network

lated to their effects on crest accuracy by means of the procedures described in this paper.

Conclusions

1. The accuracy of an impulse voltage divider system depends on the shapes of the applied wave and the response of the system to a step function. It depends primarily on the two ratios:

$$\frac{t_1}{t_r}=\frac{\text{rise time of applied wave}}{\text{rise time of voltage divider system}}$$

$$\frac{t_1}{t_2}=\frac{\text{rise time of applied wave}}{\text{fall time of applied wave}}$$

2. Curves are presented which show the relationships between these ratios and the crest accuracy.

3. Methods presented herein provide a powerful tool for determining the measuring capabilities of impulse voltage dividers and for investigating the effects of compensating features.

Appendix. Determination of Step Response

The response of nearly all impulse voltage dividers may be obtained by applying a low-voltage (100-volt) rectangular pulse to the divider and measuring the response with an oscilloscope incorporating a high-speed amplifier. The pulse should reach its crest as rapidly as possible and have negligible droop on the tail. Generally speaking, the rise time of the front should be five to ten times as fast as the expected rise time of the divider system. The pulse should last at least as long as it takes for the divider system to reach steady-state conditions. Many vacuum-tube square-wave generators meet these requirements. A type of pulse generator well suited to the purpose incorporates a transmission line charged to about 400 volts and then discharged into a load resistor through a switch with mercury wetted contacts. Such pulse generators can produce rectangular waves rising to crest in 1 to 3 mμsec (millimicroseconds) and lasting for several microseconds.

The true waveshape of the applied pulse may be determined by applying it directly to the vertical deflection plates of the oscilloscope. If the waveshape is satisfactory it may be applied to the oscilloscope through an attenuator to check the oscilloscope. The amplified wave should rise to its crest in 10 to 20 mμsec with negligible overshoot and should not droop on the tail. The sensitivity should be great enough to produce a usable deflection with an input of 0.1 volt.

The pulse should be applied with the divider and its ground and line leads in approximately the position they will occupy in service. The line lead should not be run too close to the body of the divider. The output of the divider should be fed through the appropriate length of coaxial cable to the oscilloscope to preserve the time relationships existing in actual service.

If the rise time of the divider system is short compared to the rise times of the amplifier or the applied step function pulse the indicated rise time will be too high. An approximate correction may be made, however, from the relation

$$t_r=\sqrt{(t_{r1})^2+(t_{r2})^2+(t_{r3})^2+\ldots\ldots} \qquad (3)$$

where

t_r = over-all rise time of the system
t_{r1}, t_r, t_{tr3}, etc. = rise times of the individual components of the system

References

1. **Errors in Recording Surge Voltages, P. R. Howard.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. II, Aug. 1952, pp. 371–83.

2. **Voltage Divider for Measuring Impulse Voltage on Transmission Lines, S. B. Griscom, B. L. Lloyd, A. R. Hileman.** *AIEE Transactions*, vol. 73, pt. III-A, Apr. 1954, pp. 228–35.

3. **Peak-Voltage Measurements of Standard Impulse Voltage Waves, A. Aked.** *Monograph no. 153M*, Institution of Electrical Engineers, Nov. 1955.

4. **Etude des Diviseurs de Tension en Regime Permanent, M. A. Romano.** *Revue Générale de l'Electricité*, Paris, France, vol. 65, no. 5, May 1956, pp. 289–300.

5. **The Response of Resistance Voltage Dividers to Steep-Front Impulse Waves, A. F. Rohlfs, J. S. Kresge, F. A. Fisher.** *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 634–46.

6. **Impulse Tests and Measuring Errors, N. Hyltén-Cavallius.** *ASEA Journal*, Almänna Svenska Elektriska Aktiebolaget, Vasteras, Sweden, no. 5, 1957.

7. **Transformation Calculus and Electrical Transients (book), S. Goldman.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1949, pp. 113–20.

8. **Linear Transient Analysis (book), E. Weber.** John Wiley & Sons, Inc., New York, N. Y., vol. 1, 1954, pp. 68–73.

9. **A Step-By-Step Method for Transient Analysis of Feedback Systems with One Nonlinear Element, T. M. Stout.** *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 378–90.

10. **Short-Time Spark-Over of Gaps, J. H. Hagenguth.** *Ibid.*, vol. 56, Jan. 1937, pp. 67–76.

11. **Measurement of Voltage in Dielectric Tests.** *AIEE Standard no. 4*, also *ASA Standard C68.1*, American Standards Association, New York, N. Y., 1953.

12. **Sphere-Gap Volt-Time Curves—Reference Standards for Steep-Front Measurements, J. H. Park, H. N. Cones.** *AIEE C. P. 57-215* (available on request).

13. **Impulse Voltage Measurement, F. C. Creed.** "Proceedings, High Voltage Symposium," *Publication no. 4314*, National Research Council of Canada, Ottawa, Ont., Canada, Sept. 1956.

14. **Measuring Steep-Front Impulse Voltages, H. Linck.** *Ibid.*

◆

Discussion

J. H. Park (National Bureau of Standards, Washington, D. C.): This paper, together with reference 5 of the paper, demonstrates quite conclusively the importance of the square-wave response method for investigating divider accuracy. It should, I feel, by now be quite evident that this method is to be preferred over the sustained-frequency-response method. However, it does not offer a possibility of investigating all of the short-time surge-measurement problems.

There must be some minimum divider response time that can be accurately measured with the method. This limit is apparently set by the rise time of the oscillograph amplifier which the author gives as between 10 and 20 mμsec. When using such an amplifier, any measurement of divider response time less than 0.02 μsec would be questionable. Thus for accurately measuring peak voltages occuring in less than 0.1 μsec this method would appear to be unsuitable.

The main disadvantage of relying on the square-wave-response method completely is that it gives no indication of factors such as voltage coefficient of resistance and pulse corona discharges from the divider which became troublesome at high voltage only.

N. Hyltén-Cavallius (Allmänna Svenska Elektriska Aktiebolaget, Ludvika, Sweden): May I compliment the author on a most interesting and valuable paper. I agree with it entirely but would like to discuss two points which may be of some practical importance.

The author states that a system with $\Sigma A=T=0$ will give too high measured values when measuring full impulses. This is, of course, quite true but the practical problem in general is related to:

1. The measurement of the amplitude of impulses chopped on the front down to, say, 0.4 μsec with an error less than, say, 5%.

2. The measurement of the amplitude of full impulses with front times down to, say, 0.5 μsec with an error less than 3%.

The first measurement requires a system having $\Sigma A \leq 20$ mμsec. It is possible to build such dividers for 1 megavolt and even for higher voltages but in general only if some overshoot is permitted. As an example, consider a shielded resistor divider for about 1 megavolt used in our works. It has $\Sigma A \simeq 15$ mμsec, an overshoot of about 20%, and a subsequent undershoot of about 10%. The time for the response to reach its stationary value is about 200 mμsec.

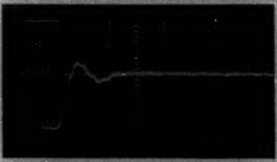


Fig. 18. Response for shielded resistor divider

The response is shown in Fig. 18 having 100-mμsec time sweep per vertical line. It can be shown by an extension of Mr. Fisher's own Fig. 7 (or preferably by Fig. 3) that the error when measuring a full impulse according to item 2 with a front time of 0.5 μsec or longer will be definitely negligible. Actually it is less than 1%.

As it is very practical to use the same divider both for full and front chopped impulses, I do not think a useful type of divider should be forbidden. A condition should, of course, be that the front of the impulse to be measured be a number of times longer than the time for the response to reach approximately its stationary value.

The author states that the 10–90% points are more useful than 30–90% for the definition of rise time. I agree entirely with this but do not understand the statement about the toe of the response and its relevance to the 10 or 30% level. The shape of the toe is quite important (we used to speak in terms of "slope at $t=0$") but it is in principle not possible to obtain any information about the shape of the toe by defining just one point on it. (The 90% level has no relevance to the toe.) It would be of interest to have some information about this.

F. A. Fisher: As Mr. Park points out, the minimum divider rise time that may be measured accurately is set by the rise time of the oscillograph amplifiers and is thus in the range of 10 to 20 mμsec for the equipment used. However, even if the rise time of the divider system is less than the rise time of the amplifier, considerable information may be gained about the shape of the actual divider step response. The indicated rise time may be approximately corrected by use of equation 3 of the paper. Thus if the indicated rise time were 16 mμsec and the amplifier rise time were 12 mμsec, the true divider rise time would be 10.5 mμsec. Oscillations on the divider step response would still be reproduced, although with lower amplitude.

Low-voltage techniques, of course, may not completely define the actual high-voltage response. A satisfactory high-voltage test procedure would be a valuable contribution to the art.

Mr. Hyltén-Cavallius' points on dividers with an oscillatory step response are well taken. I agree that such a divider would be satisfactory for measuring full waves as long as the time to crest is several times as long as the time required for the divider step response to reach its steady-state value. It should continually be kept in mind, however, that such dividers may lead to errors when measuring full waves if the front time of the full wave is short.

Mr. Hyltén-Cavallius also comments on the remarks dealing with the definitions of rise time. Actual high-voltage impulse waves often have fairly high amplitude oscillations near the start or toe of the wave which do not affect the dielectric strength of a test sample. Since the amplitudes of these oscillations are generally less than 30%, it is common practice to define the effective front time in terms of the 30% and 90% points. Similar oscillations at the beginning of a divider step-response wave would have an effect on the over-all transient response of the divider system. Defining the rise time in terms of the 10% and 90% points better reflects the effect of such oscillations on the over-all rise time.

Matrix Programming of Electronic Analog Computers

R. E. HORN
ASSOCIATE MEMBER AIEE

P. M. HONNELL
FELLOW AIEE

MATRIX PROGRAMMING of differential equations for electronic analog computers organizes the equations, thereby minimizing chances for errors; simplifies the scaling of coefficients and variables to suit the characteristics of the available machine; and in many instances permits the reduction of the number of "sign changers." The matrix method furthermore helps safeguard against the appearance of extraneous terms in the solutions of certain systems of differential equations. These very practical attributes of the matrix method of programming save computing time and equipment.

The key concept in the theory of matrix methods of programming electronic analog computers lies in the establishment of the 1-to-1 reciprocal correspondence between the matrically formulated differential equations and the synthesized electronic computer networks.[1–3] The generality of the matrix method leads not merely to one or two, but to a large group of electronic synthesis networks all yielding the solution to a given differential equation problem.

The object of this paper is to indicate by specific examples arising in practice the advantages of the matrix programming of electronic analog computers.

In the first example a wind-tunnel problem illustrates the organization, scaling of coefficients, and reduction of negative signs, thus minimizing the need for sign-changing amplifiers in a high-order system of differential equations.

This is followed by an elementary example exhibiting the care which must be taken in programming certain systems of differential equations to preclude the appearance of extraneous roots in the determinantal equation, and thereby extraneous solutions to the equations. This particularly insidious phenomenon may arise from devotion to the common "solve for the highest derivative" routine for synthesis of computer networks. These extraneous solutions are sometimes manifest as "drift" rather than in their true nature. Furthermore, the identical computer network sometimes yields "solutions" which are correct, and sometimes incorrect, depending upon the initial conditions. This is illuminated by a specific example.

Elementary Matrix Transformations

The objectives of some linear transformations of systems of linear algebraic equations and first-order differential equation systems of particular usefulness in analog computers are: 1. changing the magnitudes or signs of the coefficients of the terms in the equations, and 2. scaling in magnitude or changing the signs of the prescribed functions or dependent variables. Of course, such changes as are made must yield solutions unchanged or easily related to the original problem equations. The need for these transformations is often dictated by the purely electrical characteristics of the computer or recording equipment, such as the allowable output voltage range of the amplifiers. These elementary transformations are illustrated by the following example.

Example 1. The design of a wind-tunnel control system using analog computer techniques requires synthesis of the equations describing tunnel performance and the experimental determination of desirable controller characteristics. The particular wind-tunnel considered in this

---

Paper 58-66, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted September 3, 1957; made available for printing November 7, 1957.

R. E. HORN is with Westinghouse Electric Corporation, Baltimore, Md., and P. M. HONNELL is with Washington University, St. Louis, Mo.

example may be adequately represented by the following differential equation system, in which P represents the pressures developed as a function of the prescribed variables, A and N_0. (The differential coefficient d/dt is represented by d.)

$$(1.90+d)P_3-0.423P_4-1.59P_{ac}=-0.303A_{13} \quad (1)$$

$$-0.433P_3+(4.51+d)P_4-1.87P_{md}-0.331P_m-4.16P_p=23.7A_{30}-1.29N_0 \quad (2)$$

$$-58.1P_3+(74.9+d)P_{ac}-30.8P_p=155N_0 \quad (3)$$

$$-1.01P_4+(0.124+d)P_{md}-7.52P_s=0.302N_0 \quad (4)$$

$$(0.309+d)P_{mi}-0.159P_m=37.8N_0 \quad (5)$$

$$-0.483P_{mi}+(0.617+d)P_m=-106.A_{30}-108.A_{33} \quad (6)$$

$$-0.00514P_3-0.091P_{ac}+(2.97+d)P_p-0.973P_s=0.173.A_{13}-2.18N_0 \quad (7)$$

$$-15.6P_{md}+(16.3+d)P_s=0 \quad (8)$$

$$151.P_p-54.3P_s+10^6P_n/P_p=0 \quad (9)$$

Valve openings which determine pressure are represented by A and are assumed to be controlled by functions of the differences between desired pressures P_{ms}, P_{ss}, or pressure ratio $(P_n/P_p)_s$, and respective measured values P_m, P_s, (P_n/P_p). Controller characteristics may be represented functionally by the following equations

$$A_{33}=f_1(P_{ms}-P_m) \quad (10)$$

$$A_{30}=f_2(P_{ss}-P_s) \quad (11)$$

$$A_{13}=f_3[(P_n/P_p)_s-(P_n/P_p)] \quad (12)$$

To illustrate application of the matrix technique to this problem, the synthesis procedure for this 8th-order system is described. Inclusion of controller characteristics involves an addition to the basic matrix equation and is not essential for the purpose of the present example.

An equivalent matrix representation of the system of equations 1 through 9 is indicated in equation 13 (see below).

The problem is to adjust the coefficients to reasonable values for use with a particular computer. Since the dependent variables are not known a priori, the term reasonable initially involves a "best guess" that tends to keep both the magnitude of the coefficients and dependent variables within acceptable limits. A convenient method is to apply transformations that attempt to establish coefficients approaching unity. For the majority of electronic analog computers, the step between the final mathematical form and the synthesis then involves the introduction of a factor of 10^{-6} in all terms, thus permitting immediate interpretation in terms of computer component values expressed in microfarads and micromhos, and computer variables expressed in volts and microamperes.

In equation 13 premultipliers affecting rows 3, 8, and 9, as follows:

row 3—divide by 100
row 8—divide by 10
row 9—divide by 100

result in reasonable values for all terms except that in row 9, column 9, which remains beyond normal limits. For cases of this type application of postmultipliers affecting columns in the coefficient matrix and also the dependent variables achieves the desired result. In particular, applying the column transformation

column 9—divide by 1,000

completes the coefficient adjustment necessary in equation 13.

Treatment of coefficients having an associated negative sign conventionally requires an additional amplifier to act as a sign changer. Reduction of the number of such sign changer is of advantage from both equipment and stability standpoints. Using the matrix formulation, transformations changing sign are easily applied and the effects of each transformation relative to the reduction of minus signs is immediately evident. Although unnecessary mathematically, electronic network stability usually dictates that transformation of rows and columns are made in pairs thus maintaining principal diagonal terms with positive sign. Desirable transformations for equation 12 change the signs of:

row 1, column 1
row 4, column 4
row 6, column 6
row 7, column 7
row 9, column 9

$$\begin{bmatrix} (1.90+d) & 0.423 & 1.59 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.433 & (4.51+d) & 0 & 1.87 & 0 & 0.0331 & 4.16 & 0 & 0 \\ 0.581 & 0 & (0.749+0.01d) & 0 & 0 & 0 & 0.308 & 0 & 0 \\ 0 & 1.01 & 0 & (0.124+d) & 0 & 0 & 0 & 7.52 & 0 \\ 0 & 0 & 0 & 0 & (0.309+d) & 0.159 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.483 & (0.617+d) & 0 & 0 & 0 \\ -0.00514 & 0 & 0.091 & 0 & 0 & 0 & (2.97+d) & 0.973 & 0 \\ 0 & 0 & 0 & 1.56 & 0 & 0 & 0 & (1.63+0.1d) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.51 & 0.543 & 1 \end{bmatrix} \begin{bmatrix} -P_3 \\ P_4 \\ P_{ac} \\ -P_{md} \\ P_{mi} \\ -P_m \\ -P_p \\ P_s \\ -10^3P_n/P_p \end{bmatrix} = \begin{bmatrix} 0.303A_{13} \\ 23.7A_{30}-1.29N_0 \\ 1.55N_0 \\ -0.302N_0 \\ 37.8N_0 \\ 106A_{30}+108A_{33} \\ -0.173A_{13}+2.18N_0 \\ 0 \\ 0 \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} (1.90+d) & -0.423 & -1.59 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.433 & (4.51+d) & 0 & -1.87 & 0 & -0.0331 & -4.16 & 0 & 0 \\ -58.1 & 0 & (74.9+d) & 0 & 0 & 0 & -30.8 & 0 & 0 \\ 0 & -1.01 & 0 & (0.124+d) & 0 & 0 & 0 & -7.52 & 0 \\ 0 & 0 & 0 & 0 & (0.309+d) & -0.159 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.483 & (0.617+d) & 0 & 0 & 0 \\ -0.00514 & 0 & -0.091 & 0 & 0 & 0 & (2.97+d) & -0.973 & 0 \\ 0 & 0 & 0 & -15.6 & 0 & 0 & 0 & (16.3+d) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 151. & -54.3 & 10^6 \end{bmatrix} \begin{bmatrix} P_3 \\ P_4 \\ P_{ac} \\ P_{md} \\ P_{mi} \\ P_m \\ P_p \\ P_s \\ P_n/P_p \end{bmatrix} = \begin{bmatrix} -0.303A_{13} \\ 23.7A_{30}-1.29N_0 \\ 155.N_0 \\ 0.302N_0 \\ 37.8N_0 \\ -106.A_{30}-108.A_{33} \\ 0.173A_{13}-2.18N_0 \\ 0 \\ 0 \end{bmatrix} \quad (13)$$

The number of sign changers has now been reduced from 17 to one.

Results of the application of both scale-factor and sign-change transformations are indicated in equation 14 (see previous page), which is equivalent to equation 13 but more applicable to analog computer synthesis.

If equation 13 is represented as

$$[f_0][x_0]=[y_0] \tag{15}$$

and equation 14 as

$$[f_1][x_1]=[y_1] \tag{16}$$

equations 17 through 19 represent the elementary transformations.

$$[f_1]=\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.01 \end{bmatrix}[f_0]\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.001 \end{bmatrix} \tag{17}$$

$$[x_1]=\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1{,}000 \end{bmatrix}[x_0] \tag{18}$$

$$[y_1]=\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.01 \end{bmatrix}[y_0] \tag{19}$$

In view of their elementary character, the matrix premultipliers and postmultipliers in equations 17 through 19 need not be written down explicitly, although the information they contain, especially in the last, must be recorded.

The electronic network equation corresponding to equation 14 is indicated in equation 20, in which the coefficients now represent micromhos and microfarads, and the resultant ideograph is shown in Fig. 1.

Extraneous Solutions

The elementary solve-for-the-highest-derivative technique of computer programming, although restricted in form, usually gives correct solutions, particularly for single nth-order differential equations. In systems of differential equations, however, great care must be exercised to avoid extraneous solutions.

It is known that differentiation of equations in a system may introduce additional roots in the determinantal equation, and thus alien solutions. This constraint is easily overlooked in attempting to program a computer according to the currently accepted methods which require that each equation be written in such a manner that a derivative of a dependent variable appears to be a known function. Then, assuming the dependent variables are known, successive integrations are performed, functions combined, and eventually "fed back," completing the synthesis network. This common technique is not applicable to many systems of equations and the differentiation required in attempting to apply the method may easily yield erroneous results. Example 2 is an elementary illustration of this type of programming error.

Example 2. The following system of differential equations

$$\begin{aligned} &d^2x+d^2y-d^2z+dy+x+y=0 \\ &dy+z=0 \\ &2dx+dz+z=0 \end{aligned} \tag{21}$$

has the determinantal equation

$$\lambda(\lambda+1)(\lambda^2+1)=0 \tag{22}$$

and the general solution

$$\begin{aligned} &x=\frac{(c_3-c_2)}{2}\cos t-\frac{(c_2+c_3)}{2}\ \sin t-c_4 \\ &y=c_1e^{-t}-c_2 \sin t+c_3 \cos t+c_4 \\ &z=c_1e^{-t}+c_2 \cos t+c_3 \sin t, \end{aligned} \tag{23}$$

in which the four constants specified in equation 23 are determined by the conditions

$$\begin{aligned} &x(0)=\frac{(c_3-c_2)}{2}-c_4 \\ &y(0)=c_1+c_3+c_4 \\ &z(0)=c_1+c_2 \\ &dx(0)=-\frac{(c_2+c_3)}{2} \end{aligned} \tag{24}$$

It may appear desirable to differentiate the second and third equations of equation 21 yielding a set of equations which may then be programmed without difficulty using the conventional procedure.

The new system reads

$$\begin{aligned} &d^2x+d^2y-d^2z+dy+x+y=0 \\ &d^2y+dz=0 \\ &2d^2x+d^2z+dz=0 \end{aligned} \tag{25}$$

and has the determinantal equation

$$\lambda^3(\lambda+1)(\lambda^2+1)=0 \tag{26}$$

and the general solution

$$\begin{aligned} &x=\frac{(c_3-c_2)}{2}\cos t-\frac{(c_2+c_3)}{2}\ \sin t-(c_4+c_5)-c_6t \\ &y=c_1e^{-t}-c_2 \sin t+c_3 \cos t+c_4+c_6t \\ &z=c_1e^{-t}+c_2 \cos t+c_3 \sin t+c_5 \end{aligned} \tag{27}$$

$$\begin{bmatrix} (1.90+d) & 0.423 & 1.59 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.433 & (4.51+d) & 0 & 1.87 & 0 & 0.033 & 4.16 & 0 & 0 \\ 0.581 & 0 & (0.749+0.01d) & 0 & 0 & 0 & 0.308 & 0 & 0 \\ 0 & 1.01 & 0 & (0.124+d) & 0 & 0 & 0 & 7.52 & 0 \\ 0 & 0 & 0 & 0 & (0.309+d) & 0.159 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.483 & (0.617+d) & 0 & 0 & 0 \\ -0.005 & 0 & 0.091 & 0 & 0 & 0 & (2.97+d) & 0.973 & 0 \\ 0 & 0 & 0 & 1.56 & 0 & 0 & 0 & (1.63+0.1d) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.51 & 0.543 & 1 \end{bmatrix}\begin{bmatrix} -w_1 \\ -w_2 \\ -w_3 \\ -w_4 \\ -w_5 \\ -w_6 \\ -w_7 \\ -w_8 \\ -w_9 \end{bmatrix}=\begin{bmatrix} j_1' \\ j_2' \\ j_3' \\ j_4' \\ j_5' \\ j_6' \\ j_7' \\ j_8' \\ j_9' \end{bmatrix} \tag{20}$$

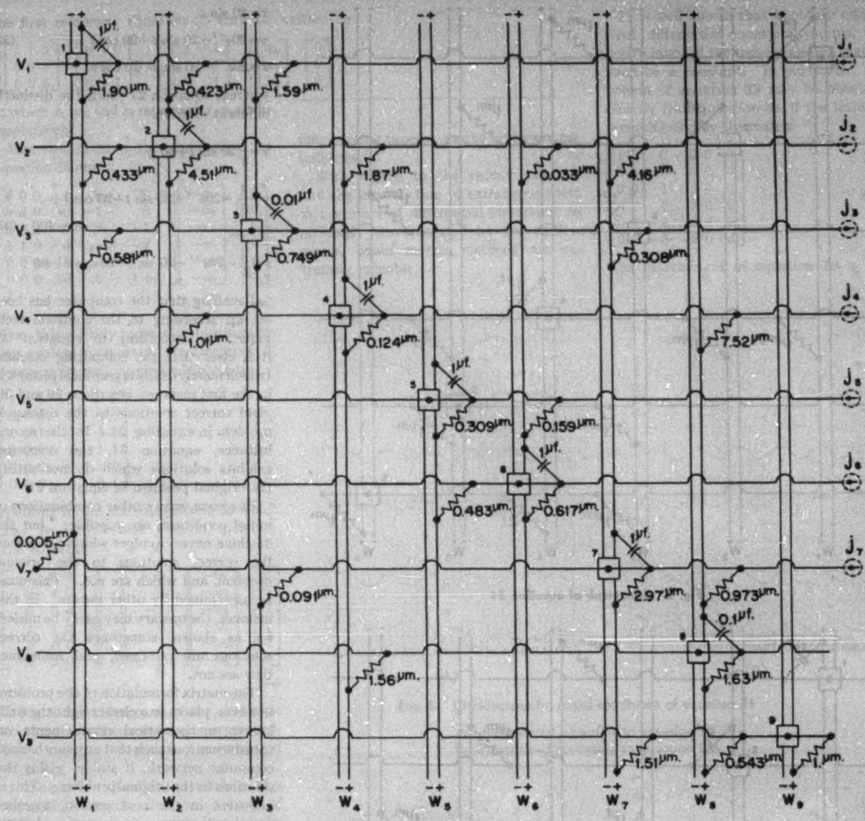

V₁
V₂
V₃
V₄
V₅
V₆
V₇
V₈
V₉
J₁
J₂
J₃
J₄
J₅
J₆
J₇
W₁
W₂
W₃
W₄
W₅
W₆
W₇
W₈
W₉
1μf.
1.90μm.
0.423μm.
1.59μm.
1μf.
0.433μm.
4.51μm.
1.87μm.
0.033μm.
4.16μm.
0.01μf
0.581μm.
0.749μm.
0.308μm.
1μf.
1.01μm.
0.124μm.
7.52μm.
1μf.
0.309μm.
0.159μm.
1μf.
0.483μm.
0.617μm.
0.005μm.
1μf.
0.091μm.
2.97μm.
0.973μm.
0.1μf.
1.56μm.
1.63μm.
1.51μm.
0.543μm.
1μm.

Fig. 1. Ideograph of equation 20

The six constants in equation 27 are related to the initial conditions by the relations

$$
\begin{aligned}
x(0) &= \frac{(c_3 - c_2)}{2} - c_4 - c_6 \\
y(0) &= c_1 + c_3 + c_4 \\
z(0) &= c_1 + c_2 + c_5 \\
dx(0) &= -\frac{(c_2 + c_3)}{2} - c_5 \\
dy(0) &= -c_1 - c_2 + c_6 \\
dz(0) &= -c_1 + c_3
\end{aligned}
\qquad (28)
$$

Comparison of the solution in equation 23 with that in equation 27 clearly indicates the appearance of extraneous constant and linear terms in the latter.

In the general situation, the solutions to the differential equation system are not known explicitly (else why use the computer?). This leads to the following interesting situation. Assume that the conventional technique has been followed and the computer set up according to the differentiated system equation 25. Now suppose the transient response to the initial conditions

$$
\begin{aligned}
x(0) &= 20 \\
y(0) &= z(0) = dx(0) = 0
\end{aligned}
\qquad (29)
$$

is desired. Then the solutions given by equations 23 and 27 both read

$$
\begin{aligned}
x &= \frac{1}{3}(20 \cos t + 40) \\
y &= \frac{1}{3}(20e^{-t} + 20 \cos t + 20 \sin t - 40) \\
z &= \frac{1}{3}(20e^{-t} - 20 \cos t + 20 \sin t),
\end{aligned}
\qquad (30)
$$

i.e., either system yields the identical solution.

Next, consider the solution for the initial conditions

$$
\begin{aligned}
x(0) &= y(0) = z(0) = 0 \\
dx(0) &= 20
\end{aligned}
\qquad (31)
$$

The solution in equation 23 to the original problem of equation 21 now yields

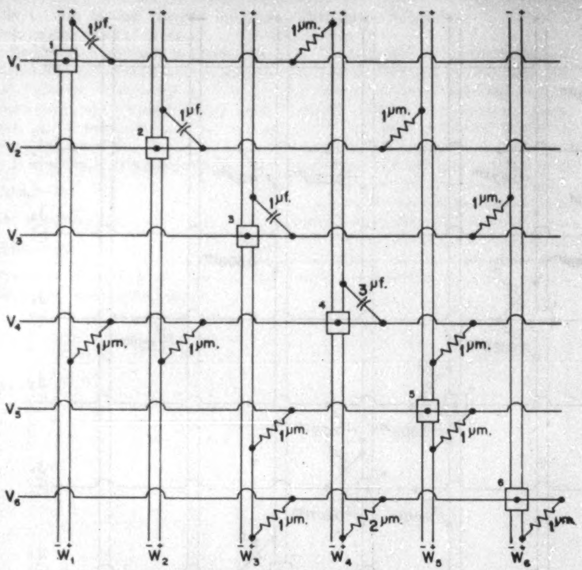


Fig. 2. Ideograph of equation 34

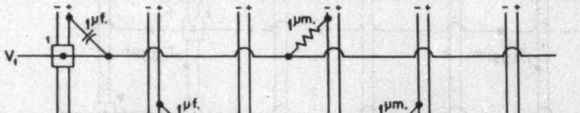


Fig. 3. Ideograph of equation 38

$$
\begin{aligned}
x &= 20 \sin t \\
y &= 20e^{-t} + 20 \sin t - 20 \cos t \qquad (32) \\
z &= 20e^{-t} - 20 \sin t - 20 \cos t
\end{aligned}
$$

whereas equation 27 yields the distinctly different solution

$$
\begin{aligned}
x &= \frac{1}{3}(20 \sin t + 40\, t) \\
y &= \frac{1}{3}(-20e^{-t} + 20 \sin t - 20 \cos t + \\
&\qquad 40 - 40t) \qquad (33) \\
z &= \frac{1}{3}(-20e^{-t} - 20 \sin t - 20 \cos t + 40)
\end{aligned}
$$

Recalling that the computer has been set up according to the common technique, i.e., according to equation 25, it is clear that the computing machine (which merely exhibits graphical plots) will in the first instance, equations 29 and 30, yield correct solutions to the ostensible problem in equation 21. In the second instance, equation 31, the computer exhibits solutions which do not satisfy the original problem in equation 21.

Of course, many other combinations of initial conditions are possible; but the machine never divulges which are indeed the correct solutions to the original problem, and which are not. This must be ascertained by other means. In this instance, the unwary may easily be misled, for as shown, sometimes the correct solutions are obtained, and sometimes they are not.

The matrix formulation of the problem, however, places in a clearer light the well-known mathematical requirements on transformations such that any synthesized computer network, if stable, yields the solutions to the original problem. This is indicated in the next section, together with oscillograms of computer solutions to the differential equation problem of equations 21 and 25 programmed by the matrix method.

Matrix Programming

The matrix programming of the differential system of equation 21 has been indicated and explained in reference 1, and one version leads to the equation

$$
\begin{bmatrix} d & 0 & 0 & -1 & 0 & 0 \\ 0 & d & 0 & 0 & -1 & 0 \\ 0 & 0 & d & 0 & 0 & -1 \\ 1 & 1 & 0 & 3d & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ -w_3 \\ -w_4 \\ -w_5 \\ -w_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (34)
$$

which is illustrated by the ideograph of Fig. 2. It therefore suffices to synthesize the differentiated system of equation 25.

According to the method outlined in the

the first reference, equation 25 may be written as

$$\left\{\begin{bmatrix}1&1&0\\0&0&0\\0&0&0\end{bmatrix}+\begin{bmatrix}0&1&0\\0&0&1\\0&0&1\end{bmatrix}D+\begin{bmatrix}1&1&-1\\0&1&0\\2&0&1\end{bmatrix}D^2\right\}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \qquad (35)$$

in which x_1, x_2, and x_3 represent x, y, and z respectively.

One first-order system equivalent to equation 35 reads

$$\begin{bmatrix}d&0&0&-1&0&0\\0&d&0&0&-1&0\\0&0&d&0&0&-1\\1&1&0&d&(1+d)&-d\\0&0&0&0&d&1\\0&0&0&2d&0&(1+d)\end{bmatrix}\begin{bmatrix}-w_1\\-w_2\\-w_3\\-w_4\\-w_5\\-w_6\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \qquad (36)$$

Since this is likely to be unstable on existing analog computers, see reference 2, equation 36 may be premultiplied by

$$\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&-1&1\\0&0&0&0&1&0\\0&0&0&-2&2&1\end{bmatrix} \qquad (37)$$

resulting in the matrix equation 38 which may be synthesized without difficulty and is stable.

$$\begin{bmatrix}d&0&0&-1&0&0\\0&d&0&0&-1&0\\0&0&d&0&0&-1\\1&1&0&3d&1&0\\0&0&0&0&d&1\\-2&-2&0&0&-2&3(1+d)\end{bmatrix}\begin{bmatrix}-w_1\\-w_2\\-w_3\\-w_4\\-w_5\\-w_6\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \qquad (38)$$

Fig. 3 is an ideograph of the electronic computer synthesis network corresponding to equation 38, and may be compared with Fig. 2.

Comparison of solutions exhibited by an analog computer for the two sets of differential equations corroborate the assertions of the previous section.

Figs. 4(A) and (B) respectively show the computer solutions to equations 21 and 25 programmed according to the matrix equations 34 and 38 for initial conditions of equation 29. As stated, the machine solutions coincide in this case.

For the initial conditions of equation 31, the oscillograms Figs. 5(A) and (B) correspond respectively to equations 32 and 33. As predicted, there is now no correlation between the respective solutions exhibited by the computer.

Résumé

The practical merits of the matrix programming of analog computers have been briefly demonstrated in this paper. The most elementary of matrix transformations are an invaluable time- and equipment-saving process, as the first example indicates.

With regard to the second example and the introduction of extraneous roots in a system of differential equations, so strikingly demonstrated by the oscillograms, again matrix methods are extremely valuable.

It is well known that to obtain equivalent differential equation systems, the determinant of the transformation matrix must be a constant. In particular, the system of equation 38 may be obtained directly from equation 34, if the latter is premultiplied by the matrix

$$\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&-1&0&d&0\\0&0&-3&-2&0&3d\end{bmatrix} \qquad (39)$$

The determinant of equation 39 is $3d^3$.

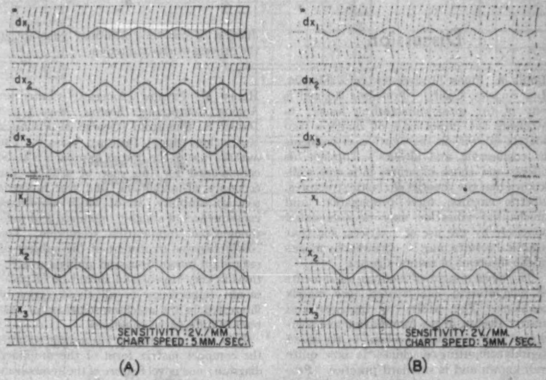


Fig. 4. Oscillograms for initial conditions of equation 31

A—Programmed according to equation 21
B—Programmed according to equation 25

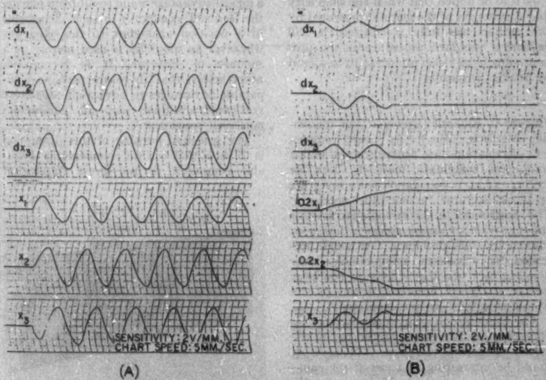


Fig. 5. Oscillograms for initial conditions of equation 31

A—Programmed according to equation 21
B—Programmed according to equation 25

Clearly this is not a constant and accounts for the extraneous λ^3 term in the determinantal equation 26, and additional terms in the solution equation 33.

In many practical problems which contain equations that have been differentiated for programming purposes, the manner of excitation of the network and initial conditions are such that the terms corresponding to the extraneous roots should have zero coefficients. However, whether due to initial or residual charges on the capacitors, grid current, internal parasitic coupling, or inexact matching of components, a small constant or linear term may be excited in addition to the correct solution response. The presence of such linear terms in the exhibited solution may then be interpreted as amplifier "drift," rather than actually as an incorrect synthesis network. This is the situation which may easily arise as indicated by the second example.

Finally, it should be clear that the application of matrix methods to analog computer programming are of far wider scope and generality than can be indicated in this brief introduction. But the practicality of the matrix method is clearly established. If not employed, the most economic operational use of the machine is not realized. This is the antithesis of the *raison d'être* for the computing machine in the first place!

References

1. Matrices in Analogue Mathematical Machines, Pierre M. Honnell, Robert E. Horn. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 260, no. 3, Sept. 1955, pp. 193–207.

2. Matrices in Electronic Differential Analyzers, Pierre M. Honnell, Robert E. Horn. *Proceedings*, International Analogy Computation Meeting, Brussels, Belgium, 1956, pp. 217–21.

3. Analogue Computer Synthesis and Error Matrices, P. M. Honnell, R. E. Horn. *AIEE Transactions*, vol. 75, pt. I, Mar. 1956, pp. 26–32.

◆

Discussion

Louis A. Pipes (University of California, Los Angeles, Calif.): Professor Honnell is one of the most outstanding leaders in stressing the importance of matrices and matrix calculus in the solution of problems in engineering and physics. Through his efforts and those of others it is now well-known that the theories of elastic structures, electric circuits, wave propagation, and mechanical vibration may be formulated concisely by the use of matrices, and that practical results may be obtained by means of the theorems of matrix algebra.

The development of modern instruments of computation has created new matrix problems. The utility of matrices and matrix algebra in the formulation and solution of many physical problems by the use of digital computing machines is now quite well known and is standard practice. Professor Honnell and Dr. Horn are to be congratulated for demonstrating the power and elegance of matrix methods in systematizing the programming of physical problems in analog computing machines.

It is to be hoped that when the simplicity and conciseness of the method presented by the authors are fully recognized, it will become a standard procedure in the programming of physical problems to be solved by analog computers.

Vincent C. Rideout (University of Wisconsin, Madison, Wis.): As more complex sets of equations are attempted on analog computers, systematic methods, both for scaling and for manipulation prior to setting up block diagrams, become more necessary. The matrix methods introduced in this paper provide a valuable guide to those confronted by such problems. Although presented for equations representing linear systems, they are also of help in nonlinear system study, if linearized forms of the system equations are first examined.

The introduction of extraneous roots by differentiation is clearly pointed out in the second example of the paper. However, it would be interesting to know if the equations of this example could come directly from a physical system. Also, this set may be shown to require only four arbitrary constants by the fact that it can be reduced to a third-order equation in y with the introduction of only one arbitrary constant in the process of this reduction.

Henry M. Paynter (American Center for Analog Computing, Boston, Mass.) and **Daniel H. Sheingold** (George A. Philbrick Researches, Inc., Boston, Mass.): The authors' use of the matrix approach is to be commended for the ease of scaling, the economy of amplifiers, and the inherent solution stability, all of which are direct consequences of a highly organized and well-thought-out formulation technique. The comprehensive treatment of the subject of redundant integrations and ambiguous solutions is very timely and should be of paramount importance to all those seriously interested in the machine solutions of systems of differential equations.

It is also particularly interesting to note the compact matrix form of the amplifier diagram; one novel feature of the symbolism is the use of conductance measures rather than resistances which results in a direct relationship between coefficients in the equations and physical parameters in the machine. However, we might suggest that a logical advantage could result from placing the feedback conductor in such ideographs in a position analogous to the capacitor whenever an amplifier is used merely as a summer or inverter. It is recognized, nevertheless, that there are perhaps comparable advantages to having all capacitors northwest/southeast and all conductors northeast/southwest.

The principal purpose of this discussion is to show that certain commercially available computers permit direct translation from the original matrix form (and, of course, from the original equations) to machine interconnections and parametric settings. Such machines are composed of modular elements which are called, for convenience, Universal Linear Operators. Fig. 6 shows the front panel of one modular element, and Fig. 7 shows a bank of 12 such elements. The basic equation, which completely describes the performance of each element is

$$e=e_0+10^m p^{-n}\sum_{i=1}^{4} a_i e_i \qquad (40)$$

where

$m=0, 1, 2, 3$

$n=0, 1$

$$1/p=\int_0^t (\ )dt \qquad (41)$$

The signal e_0 is adjustable to any positive or negative constant voltage within the output range of the element in steps of 0.1 volt;

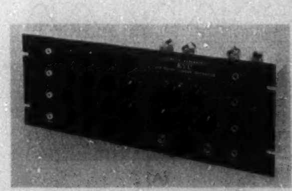


Fig. 6. Front view of a Universal Linear Operator showing decade parameter settings and external connections

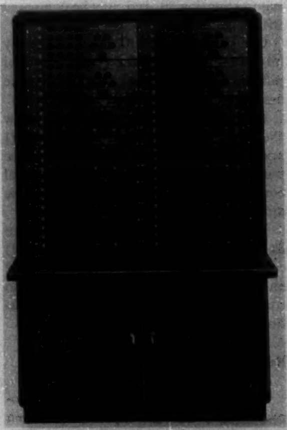
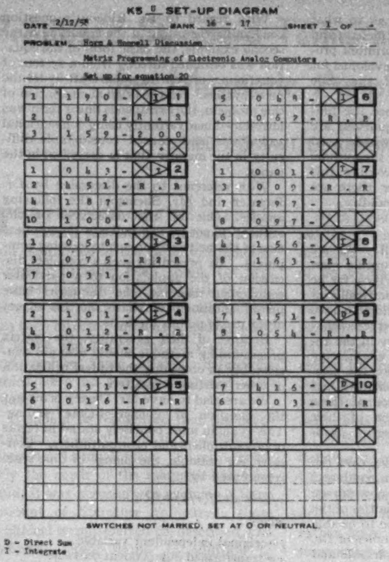


Fig. 7. A bank of 12 Universal Linear Operators



K5 U SET-UP DIAGRAM

DATE 2/12/58 BANK 16 - 17 SHEET 1 OF -

PROBLEM Horn & Honnell Discussion

Matrix Programming of Electronic Analog Computors

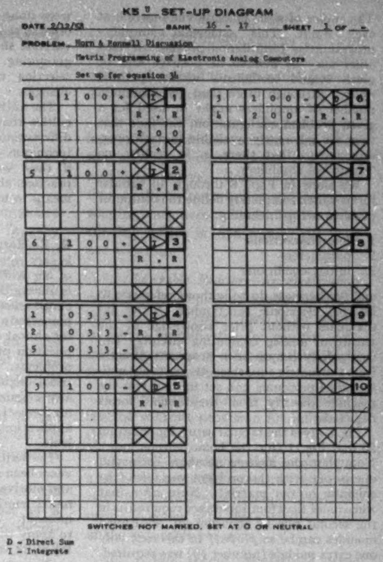
Set up for equation 20

1		1	9	0	-	☒	I▷	1	5		0	4	8	-	☒	I▷	6
2		0	4	2	-	R	.	R	6		0	6	2	-	R	.	R
3		1	5	9	-	2	0	0									
						☒	+	☒							☒		☒
1		0	4	3	-	☒	I▷	2	1		0	0	1	+	☒	I▷	7
2		4	5	1	-	R	.	R	3		0	0	2	-	R	.	R
4		1	8	7	-				7		2	9	7	-			
10		1	0	0	+	☒		☒	8		0	9	7	-	☒		☒
1		0	5	8	-	☒	I▷	3	4		1	5	6	-	☒	I▷	8
3		0	7	5	-	R	2	R	8		1	6	3	-	R	1	R
7		0	3	1	-												
						☒		☒							☒		☒
2		1	0	1	-	☒	I▷	4	7		1	5	1	-	☒	D▷	9
4		0	1	2	-	R	.	R	8		0	5	4	-	R	.	R
8		7	5	2	-												
						☒		☒							☒		☒
5		0	3	1	-	☒	I▷	5	7		4	1	6	-	☒	D▷	10
6		0	1	6	-	R	.	R	6		0	0	3	-	R	.	R
						☒		☒							☒		☒

SWITCHES NOT MARKED, SET AT 0 OR NEUTRAL.

D - Direct Sum
I - Integrate

Fig. 8 (left). Tabular setup for equation 20



K5 U SET-UP DIAGRAM

DATE 2/12/58 BANK 16 - 17 SHEET 1 OF -

PROBLEM Horn & Honnell Discussion

Matrix Programming of Electronic Analog Computors

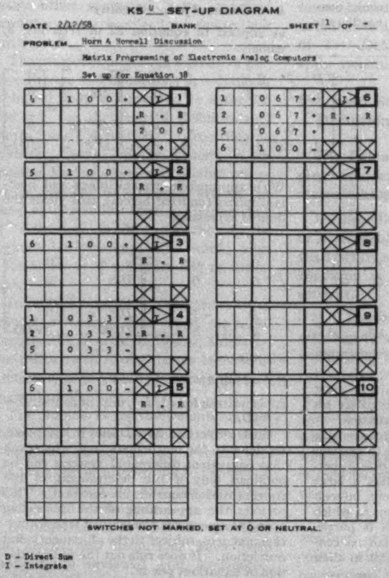
Set up for equation 34

4		1	0	0	+	☒	I▷	1	3		1	0	0	-	☒	D▷	6
						R	.	R	4		2	0	0	-	R	.	R
						2	0	0									
						☒	+	☒							☒		☒
5		1	0	0	+	☒	I▷	2							☒	▷	7
						R	.	R									
						☒		☒							☒		☒
6		1	0	0	+	☒	I▷	3							☒	▷	8
						R	.	R									
						☒		☒							☒		☒
1		0	3	3	-	☒	I▷	4							☒	▷	9
2		0	3	3	-	R	.	R									
5		0	3	3	-												
						☒		☒							☒		☒
3		1	0	0	-	☒	D▷	5							☒	▷	10
						R	.	R									
						☒		☒							☒		☒

SWITCHES NOT MARKED, SET AT 0 OR NEUTRAL.

D - Direct Sum
I - Integrate

Fig. 9 (right). Tabular setup for equation 34



K5 U SET-UP DIAGRAM

DATE 2/12/58 BANK SHEET 1 OF -

PROBLEM Horn & Honnell Discussion

Matrix Programming of Electronic Analog Computors

Set up for Equation 38

4		1	0	0	+	☒	I▷	1	1		0	6	7	+	☒	I▷	6
						R	.	R	2		0	6	7	+	R	.	R
						2	0	0	5		0	6	7	+			
						☒	+	☒	6		1	0	0	-	☒		☒
5		1	0	0	+	☒	I▷	2							☒	▷	7
						R	.	R									
						☒		☒							☒		☒
6		1	0	0	+	☒	I▷	3							☒	▷	8
						R	.	R									
						☒		☒							☒		☒
1		0	3	3	-	☒	I▷	4							☒	▷	9
2		0	3	3	-	R	.	R									
5		0	3	3	-												
						☒		☒							☒		☒
6		1	0	0	-	☒	I▷	5							☒	▷	10
						R	.	R									
						☒		☒							☒		☒

SWITCHES NOT MARKED, SET AT 0 OR NEUTRAL.

D - Direct Sum
I - Integrate

Fig. 10 (left). Tabular setup for equation 38

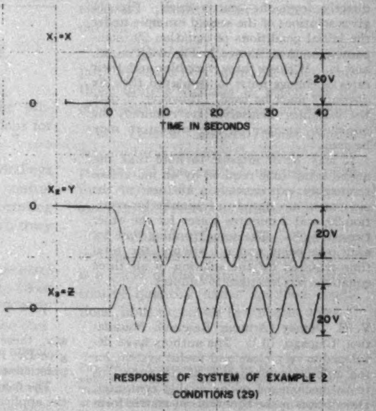


Fig. 11 (right). Solutions of equation 34 for x_1, x_2, and x_3

and a_i is adjustable to any positive or negative value from 0 to 11.10 in steps of 0.01 by means of 3-digit decade settings. It is seen immediately that this operator may function either as an integrator ($n=1$) or as a summer ($n=0$), and that the m parameter permits, in effect, a choice among four values of feedback resistance or capacitance. The voltage e_0 then functions either as a directly additive constant in the case of simple summation, or as an initial condition in the case of integration.

The only fundamental distinction affecting problem formulation between the operators described in the foregoing and operational summing points in the authors' matrix diagrams is in the location of the inverting amplifiers. The operational elements described here have outputs which are of one sense only, i.e., "direct," or positive.

The choice of signs is made via a switch at each input. This somewhat simplifies programming formulation because all variables may be defined as positive and be given sensed coefficients where operated upon. Although more than the minimum number of amplifiers might be used for a given problem configuration, the fact that they are inherently available in each operator means that there is, in reality, no "waste" of amplifiers.

We show in Figs. 8 through 10 tabular forms which completely define the computor problem setup including:

1. interconnections
2. parameters
3. initial conditions

and furthermore do so without reference to internal electronic functioning. Thus, individuals without prior knowledge of the nature of analog computing circuitry can successfully set up such programs directly from the matrix equations, and entirely nontechnical personnel can set up and run the machine directly from the tabular forms depicted.

Fig. 8 gives the tabular setup for equation 20 and Fig. 1 (for the homogeneous case). Note that the module numbers correspnd one to one with the problem variables (the columns in the matrix). Note also that when more than four inputs are required as in the second row of equation 20, additional modules can be employed; in this case only one extra module (number *10*) was required.

Figs. 9 and 10 correspond to equations 34 and 38 and Figs. 2 and 3 respectively. Note again that the parameters can be set directly from the matrix itself. Fig. 11 gives solutions of the second example under the initial conditions of equation 29, comparable to those shown in Fig. 4 for x_1, x_2, and x_3. If these three variables and their rates of change must be displayed then six operators must be used. However, if only the variables themselves are required, the minimum number of four operators may be used.

It may be of interest to note that the actual setup time requred by an individual reasonably experienced in the use of the equipment amounted to 4 minutes for equation 20, and 2 minutes apiece for the other two examples. The actual running time for each case being under 1 minute, the total time required for the solution of all three equations was less than 15 minutes.

V. H. Disney (Armour Research Foundation, Chicago, Ill.): The authors have developed a very clear and useful system for the orderly manipulation of the equations to be mechanized on an analog computer. Description of the equations in matrix form permits an easily checked layout of the programming connections which should do much to reduce the possibility of error. It should be noted, however, that only the scales of coefficients and dependent variables are changed by the procedure described. It is suggested that this system be extended to include change of time scale as well. If this were done the question arises as to what order of scale change in time, coefficient, and variable should be made for optimum programming. Guides might be developed for selecting reasonable values of the scale changes.

The authors have also described the difficulties that can arise when equations are differentiated. This is not a widely used technique. I should like to ask the authors if they would recommend avoiding this practice altogether, or whether there are means to test the solutions for validity.

R. E. Horn and P. M. Honnell: We are keenly appreciative of the favorable view of our work expressed by Dr. Pipes, for he is among the handful of mathematically inclined scientists in the entire world who have pioneered and steadfastly pursued researches for almost a quarter of a century in the application of matrix methods to engineering problems. This was long before the electrical engineering profession generally had any cognizance of the existence of these mathematic methods, much less of their application to electical engineering problems.

The writings of Dr. Pipes have, over the years been a source of instruction, guidance, and motivation to all of us who now pursue matric methods in our own specialties of the moment. In particular, we hope to be able to apply the results of the researches of Dr. Pipes, in the difficult area of variable and nonlinear systems, to the electronic computing techniques we describe; needless to say, these are matric methods! It must give Dr. Pipes great satisfactoin to see his prescience so abundantly substantiated.

The comments of Dr. Rideout concerning the application of these methods to systems of nonlinear equations are particularly pertinent. It is in these areas that a large percentage of problems of current interest lie—complex systems of equations in which the systematic matrix approach is particularly helpful. The authors plan to consider these aspects in more detail in their future publications.

With regard to Dr. Rideout's question concerning the origin of the equation in the second example, the basic form of the equation was obtained directly from a system of equations representing an aircraft hydraulic servo system. In the original simulation of this problem, the extraneous linear term was introduced inadvertently using conventional techniques and caused considerable difficulty in the manner that is described in the paper.

With reference to the discussion by Dr. Paynter and Mr. Sheingold the following applies. One of the objectives of the authors' paper was to indicate the application of the matrix approach to existing computers. By presenting such a complete discussion of the application to a particular commercial machine, the discussors have placed additional emphasis on this aspect of the use of matrix methods.

Because of their generality, the matrix programming methods should also aid considerably in obtaining the fullest utilization of other existing computers.

We are glad to have Mr. Disney bring out the question of change-of-time scaling. This is again most concisely stated in terms of premultipliers and postmultipliers. Consider, for example, the change in time scale represented by

$\tau/k=t$ and $f(\tau/k)=f(t)$,

where

t = original independent variable
τ = transformed independent variable
k = constant of proportionality

as applied to equation 22 of reference 3, namely

$$\begin{bmatrix} D & -I & 0 & \ldots & 0 & 0 \\ 0 & D & -I & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & D & -I \\ A_1 & A_2 & A_3 & \ldots & A_n & A_{n+1} \end{bmatrix} \begin{bmatrix} x \\ Dx \\ . \\ D^{n-1}x \\ D^n x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ f(t) \end{bmatrix}$$

With appropriate premultipliers and postmultipliers (omitted herein), this yields the desired expression

$$\begin{bmatrix} D_\tau & -I & 0 & \ldots & 0 & 0 \\ 0 & D_\tau & -I & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & D_\tau & -I \\ A_1 & kA_2 & k^2A_3 & \ldots & k^{n-1}A_n & k^nA_{n+1} \end{bmatrix} \begin{bmatrix} x \\ D_\tau x \\ . \\ D_\tau{}^{n-1}x \\ D_\tau{}^n x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ f(\tau/k) \end{bmatrix}$$

wherein

$D_\tau{}^n=I\times d_\tau{}^n$,

$d_\tau{}^n$ is written for $d^n/d\tau^n$, and, of course, $D=(1/k)D_\tau$.

With respect to Mr. Disney's last point, we defer to the mathematicians, who teach that equivalent differential systems can be obtained only if the determinant of the transformation matrices are constant. This permits the appearance of the derivative operator among the elements of transformation matrices, subject to the aforementioned restriction. It does rule out the differentiation of equations per se.

Vol.
Pt. 1
1958

Proposed Standard Terms and Definitions for Magnetic Amplifiers

AIEE COMMITTEE REPORT

THE PURPOSE OF THIS PAPER is to review the present status of terms and definitions in the rapidly expanding magnetic-amplifier field. Since the first progress report of the Magnetic Amplifiers Subcommittee in 1950[1] much effort has been directed towards preparing a list of terms and definitions to supplement and replace those original definitions which were no longer considered adequate.

It should be recognized that the preparation of a list of useful terms and definitions is not a small task. To serve properly the needs of the magnetic-amplifier field, it was necessary to examine a large amount of literature to extract those technical terms referred to, described, and defined. These terms, in addition to many other suggested terms, required careful scrutiny in order to evaluate their suitability.

The project of preparing a group of preferred terms and definitions in a dynamic technical field is hampered to some extent by the dictionary's definitions of "term" and "definition." According to Webster, "term" means a word or expression having a precisely limited meaning or peculiar to a science, art, or the like; as a technical term. The meaning of "definition" is the explanation of the meaning or meanings of a word; also, a formulation of such meaning or meanings.

In a technical field the existence of several meanings of a word may frequently lead to utter confusion; similarly, the existance of several terms to describe some unique or special phenomenon tends to create difficulties.

The problem facing the Definitions Subcommittee of the Magnetic Amplifiers Committee during the last several years has been one of attempting to provide in each case one term and one definition which would be adequate.

During this time, a list of more than 100 terms and definitions have been thoughtfully reviewed and considered, and this progress report is a means of acquainting persons in this field with a partial list of terms and definitions which have been agreed upon by the subcommittee.

As a result of numerous committee meetings, surveys, ballots, and lengthy discussions, there is at this time a group of 38 terms and definitions which the Definitions Subcommittee and the Magnetic Amplifiers Committee have approved. These terms and definitions have been recently reviewed and accepted by the Standards and Test Codes Subcommittee of the Magnetic Amplifiers Committee. This group of terms and definitions has been transmitted to the AIEE Standards Committee with the endorsement of the Magnetic Amplifiers Committee.

In addition to the group of terms and definitions recommended for use is a rather large group of terms and definitions which are being considered for future recommendation. Though the magnetic amplifier is a static device, the activities of the Definitions Subcommittee are truly dynamic.

Proposed Standard Terms and Definitions

The following list contains the proposed standard terms and definitions for magnetic amplifiers:

1. *Bias Windings*—The bias windings of a saturable reactor are those control windings by means of which the operating condition is translated by an arbitrary amount.

2. *Control Windings*—The control windings of a saturable reactor are those windings by means of which control magnetomotive forces are applied to the core.

3. *Feedback* (in a magnetic amplifier)—A circuit connection by means of which an additional magnetomotive force, which is a function of the output quantity, is used to influence the operating condition.

4. *Feedback Windings*—The feedback windings of a saturable reactor are those control windings to which the feedback connections are made.

5. *Magnetic Amplifier*—A device using saturable reactors either alone or in combination with other circuit elements to secure amplification or control.

6. *Output Windings*—The output windings of a saturable reactor are those windings other than feedback associated with the load and through which power is delivered to the load.

7. *Power Windings*—The power windings of a saturable reactor are those windings to which the power is supplied by a local source. Commonly the functions of the output winding and the power windings are accomplished by the same winding, which is then termed the "output winding."

8. *Saturable Reactor*—An electromagnetic device, employing one or more nonlinear magnetic cores; used in a-c circuits to secure amplification or control, commonly by means of a d-c signal which influences the nonlinearity.

9. *Saturating Reactor*—A nonlinear reactor without provision for influencing the nonlinearity.

10. *Self-Saturation*—Self-saturation in a magnetic amplifier refers to the connection of half-wave rectifying circuit elements in series with the output windings of the saturable reactors.

11. *Signal (input, control)*—A signal is an independent input variable. It is applied to the magnetic amplifier as an electrical quantity which controls the output.

12. *Signal Windings*—The signal (input) windings of a saturable reactor are those control windings to which the independent variables (signals) are applied.

13. *Control Ampere-Turns*—Control ampere-turns expresses the magnitude and polarity of the control magnetomotive force required for operation of a magnetic amplifier at a specified output.

14. *Control Characteristic*—The control characteristic of a magnetic amplifier is a curve of the output quantity versus control quantity under specified conditions, both expressed in suitable units.

Paper 58-175, recommended by the Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted November 4, 1957; made available for printing November 13, 1957.

Personnel of the Definitions Subcommittee of the Magnetic Amplifiers Committee: D. Feldman, *chairman*, A. B. Haines, P. W. Barnhart, T. G. Blanchard, V. B. Boros, H. F. Eckbels, L. A. Finzi, A. E. Fitzgerald, W. A. Geyger, A. Krausz, R. C. Lewis, H. W. Lord, R. E. Morgan, P. Siskind, D. H. Smith, H. F. Storm, and G. L. Sullivan. Space does not permit listing the names of the many other persons who have contributed their time and effort to the preparation of this paper: the contributions of past members of the Magnetic Amplifiers Committee and its various subcommittees are gratefully acknowledged.

15. *Cross-Conduction*—Cross conduction in a reversible magnetic amplifier is the simultaneous conduction or gating of half-wave elements connected in a mesh containing the a-c supply voltage but not the load.

16. *Drift*—Drift in a magnetic amplifier is a change in the control characteristic due to unassignable causes over a specified time interval.

17. *Dynamic Coercive Force*—That value of applied magnetomotive force at which a major dynamic hysteresis loop crosses the abcissa axis (has zero value of magnetization).

18. *Dynamic Hysteresis Loop* (*A-C Hysteresis Loop, Flux-Current Loop, Flux-Ampere-Turn Loop*)—The dynamic hysteresis loop of a core is the curve of magnetization versus applied magnetomotive force per unit length obtained when the magnetic material is cyclically magnetized at some specified rate. The magnetization at any instant is the average flux density over the entire core section. The applied magnetomotive force per unit length at any instant is that required to provide the flux and rate of change of flux implicit in the test or operating specifications.

19. *Dynamic Magnetizing Force*—The applied magnetomotive force per unit length required to provide the flux and the rate of change of flux specified by the conditions for which a dynamic hysteresis loop is determined.

20. *Dynamic Permeability*—The permeability determined by the slope of a line joining the tips of a dynamic hysteresis loop under specified conditions.

21. *Firing*—Firing in a magnetic amplifier is the transition from the unsaturated to the saturated state of the saturable reactor during the conducting or gating alternation. Firing is also used as an adjective modifying phase or time to designate when firing (n) occurs.

22. *Ferroresonance* (See also Nonlinear Resonance)—If the dissipation of a passive electric network, containing at least one capacitor and at least one nonlinear inductor, is sufficiently low, the curve of response (e.g., rms value) of the network to an electric driving signal is double-valued over a certain range, when plotted against the independent variable of driving force, or frequency, or one of the network parameters. Such a network is said to be in ferroresonance, if the operating point is on the upper leg of this double-valued response curve, and if the lowest-frequency component of the response is of the same frequency as the fundamental frequency of the driving force.

23. *Ideal Saturable Core*—A core displaying idealized magnetic characteristics, i.e., a magnetization curve represented by two straight lines, having u=zero above the knee and u=infinity below the knee, and having no losses.

24. *Ideal Saturable Reactor*—A saturable reactor employing ideal cores, having perfect electromagnetic coupling among all windings on a common core and characterized by absence of losses and capacitances.

25. *Ideal Transformer*—A transformer employing ideal core or cores operating in the unsaturated region of the core characteristic, having perfect coupling between windings and further characterized by absence of winding resistances, losses, and capacitances.

26. *Impeded Harmonic Operation* (*Constrained Magnetization, Forced Magnetization*)—Impeded harmonic operation denotes that type of operation which takes place in a magnetic amplifier in which the impedance of the control circuit and any circuit closely coupled to it is so great as to substantially prevent the flow of all harmonic currents in such circuits.

27. *Magnetic Flip-Flop*—A bistable circuit employing one or more magnetic amplifiers so arranged as to have two discrete levels of output which may be obtained by adjusting the control voltage or current.

28. *Magnettor*—A second harmonic-type magnetic modulator.

29. *Nominal Control Ampere-Turns*—The difference between the control ampere-turns for minimum output and for rated output is the nominal control ampere-turns difference or, briefly, control ampere-turns.

30. *Nonlinear Resonance* (See also Ferroresonance)—Provided the dissipation of the system is low, the curve of the response (e.g., rms value) to a driving force, of a double energy-storage, passive system containing at least one nonlinear energy storage element, may be double-valued over a certain range, when plotted against the independent variable of driving force, or frequency, or one of the system parameters. Such a system is said to be in nonlinear resonance, if the operating point is on the upper leg of this double-valued response curve, and if the lowest-frequency component of the response is of the same frequency as the fundamental frequency of the driving force. If the circuit parameters or the excitation or both preclude the existence of a double-valued response, the system is said to be in nonlinear resonance if the operating point is in that range of the curve of response (e.g., rms value) versus driving force, in which the slope of the curve is greater than the slope in the vicinity of the origin.

31. *Postfiring*—Postfiring refers to the interval during the gating or output alternation following firing. It is used as an adjective to describe phenomena occuring during this interval.

32. *Prefiring*—Prefiring refers to the interval during the output or gating alternation preceding firing. It is used as an adjective to indicate phenomena occurring during this interval.

33. *Rectifier, Complementary*—Complementary rectifiers are those half-wave rectifying circuit elements in the output circuit of a magnetic amplifier which are not self-saturating rectifiers.

34. *Rectifier, Self-Saturating*—Self-saturating rectifiers are those half-wave rectifying circuit elements connected in series with output windings of a saturable reactor in the self-saturating magnetic amplifier circuit.

35. *Response Time*—The response time of a magnetic amplifier is that period of time required for a given change of the output quantity following a step change of the input quantity. This change of the output quantity shall be 63% of the total change unless otherwise specified. The "response time" specified shall be the maximum which exists for any condition within the rating (e.g., such effects as growth or decay, time phase of signal application, or temperature).

36. *Saturable Transformer*—A saturable reactor in which both output windings and power windings are employed, commonly to secure voltage transformation or circuit isolation from the a-c supply.

37. *Unimpeded Harmonic Operation* (*Natural Magnetization, Free Magnetization*)—Unimpeded harmonic operation denotes that type of operation which takes place in a magnetic amplifier in which the impedance of the control circuit or any circuit closely coupled to it is so small as to permit substantially unimpeded flow of all harmonic currents in such a circuit.

38. *Environmental Sensitivity*—Environmental Sensitivity in a magnetic amplifier is the change in the control char-

acteristic due to specified changes in environmental conditions. These include changes in ambient temperature, supply voltage, supply frequency, and other specified changes.

Reference

1. Progress Report of the AIEE Magnetic Amplifier Subcommittee, AIEE Committee Report. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 445–50.

◆

Discussion

H. F. Storm (General Electric Company, Schenectady, N. Y.): While most of the terms under discussion were in common usage many years ago, a better understanding of magnetic-amplifier function gained since then has led to changes in terminology. In the belief that a proposed standard should reflect present-day terminology, and serve as a guide for future writing rather than as a dictionary of past terms, I have prepared the following comments on some of the terms under consideration. The various terms in the following are identified by the numbers used in the paper.

6. *Output Windings*—This term appears to have been virtually discarded. For instance, at the AIEE–Institute of Radio Engineers Conference on Magnetic Amplifiers in 1957 I did not find a single paper where the term "output windings" was used. Likewise, a check through a recent volume of *Communication and Electronics*,[1] did not reveal a single instance where the term "output windings" was used. My suggestion is that the term be removed from the proposed standard under discussion. As a replacement, "gate windings" is recommended. The use of this term is widespread, and it has the further advantage of conciseness, as will be shown. The terms "gate," "gate circuit," etc., are also well established in other areas such as in connection with electronics and transistors e.g., see references 2 and 3 of this discussion.

As far as the term "gate" in magnetic amplifier work is concerned, its popularity may be gauged from the fact that "gate windings" is used in seven out of the eight magnetic-amplifier papers published in the aforementioned issue of "*Communication and Electronics*,[1] and many other examples could be listed. The spread of this term through industry across the country is indicated by the following representative list of authors of these seven papers, and their affiliations: Andress Kernick, H. A. Perkins, Jr., Westinghouse Electric Corporation, Pittsburgh, Pa.; T. T. Kadota, and H. C. Bourne, University of California, Berkeley, Calif.; W. R. Seegmiller, General Electric Company, Schenectady, N. Y.; E. W. Manteuffel, and R. O. McCary, General Electric Company, Ithaca, N. Y.; Robert M. Hubbard, Boeing Airplane Company, Seattle, Wash., R. J. Distler, and C. Sturzenbecker, Bell Telephone Laboratories, Inc., Whippany, N. J.; and H. C. Bourne, Jr., and D. Nitzan, University of California, Berkeley, Calif. The eighth paper on the subject, by R. A. Raber, does not discuss the "gate windings" and hence no term for these windings appears.

The term "gate windings" is derived from the manner in which these windings function: essentially, they either block current flow or they freely admit current flow. Similar action is encountered in many electronic and transistor circuits; these circuits are called "gate circuits" in the literature.[2,3]

The concisiveness of the term "gate" is still more apparent when it is used in conjunction with other terms. For instance, compare "voltage across the output windings" with the simple "gate voltage."

In view of the foregoing, my recommendation is to replace the presently proposed "output windings" by "gate windings," in accordance with the following definition:

The gate windings of a saturable reactor are those windings, other than feedback windings, which cyclicly permit the flow of load current.

7. *Power Windings*—In the paper, the "power windings" are defined as those". . . to which the power is supplied by a local source." I am at a loss to visualize the meaning and significance of a "local source," and feel that the definition of "power windings" should be suspended until the meaning of "local source" is established. Furthermore, it would be helpful to have described an example for the difference between the terms numbered 6 and 7.

21. *Firing*—Although in quite common usage, the term "firing" has the following drawbacks: first, nothing really fires. Second, the term may mislead the uninitiated reader to assume a tube in use somewhere, an allegation which most magnetic amplifier engineers would rather avoid. My suggestion is to replace "firing" by "saturating," with the following definition:

Saturating in a magnetic amplifier is the transition from the unsaturated to the saturated state of the saturable reactor. Saturating is also used as an adjective modifying phase or time to designate when saturating occurs.

28. *Magnettor*—This term has been used in papers on the subject only once or twice, and I have never heard it used anywhere else. The term should be dropped.

31. *Postfiring*—For reasons similar to those given for item 21, it is recommended that "postfiring" be replaced by "saturating interval."

32. *Prefiring*—For reasons similar to those given for item 21 it is recommended that "prefiring" be replaced by "presaturating."

34. *Rectifier, Self-Saturating*—Although this term is in somewhat common usage, it is cumbersome and perhaps misleading. My suggestion is to call these rectifiers "gate rectifiers," and that they be defined as follows:

Gate rectifiers are those half-wave rectifying circuit elements connected in series with the gate windings of a saturable reactor in the self-saturating magnetic amplifier circuit.

36. *Saturating Transformer*—The definition of this term is based among other things on "power windings." This latter term, however, is in need for clarification (see item 7); hence, the term "saturating transformer" should be held in abeyance pending clarification of item 7.

References

1. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 504–08; 515–20; 526–29; 562–66; 578–88; 602–06; 646–55.

2. Transistor Circuit Engineering (book), R. F. Shea. John Wiley & Sons, Inc., New York, N. Y., 1957, p. 317.

3. Electronic Analog Computers (book), Granino A. Korn, Theresa M. Korn. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, p. 272.

R. E. Anderson (General Electric Company Schenectady, N. Y.): The standards proposed by the committee should become most helpful in establishing consistent terminology in the magnetic amplifier field and thus aid in the communication and understanding of this technology. This discussion, which will be directed toward items 16 and 38, entitled drift and environmental sensitivity, is intended to cover certain details regarding the nature and effects of these particular items and thus provide insight to the possible uses and implications of the definitions. The definitions are concerned with the changes or shifts which may alter the zero or quiescent point of a magnetic amplifier. Specifically, these shifts, which may be viewed as an axis translation within the amplifier, are those which occur after the amplifier has been zeroed, i.e., adjusted to obtain a suitable quiescent point, or after it has been calibrated. Thus, gain shifts or variations dependent upon signal level are excluded from this discussion. In this way, the phenomena under consideration may be represented by an additional signal

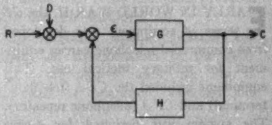


Fig. 1. Amplifier with magnetic feedback

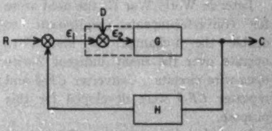


Fig. 2. Amplifier with electric feedback

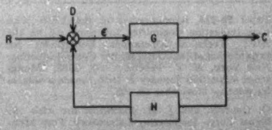


Fig. 3. Amplifier showing equivalent drift

or output of constant magnitude, which may appear in a manner independent of the signal or the loading. In the definition now proposed, we find that two types of zero shift are defined as opposed to one drift suggested by many of the previous standards. For instance, the progress report of the subcommittee[1] has led many of us to consider as drift, changes arising from both environment and unassigned causes.

NOMENCLATURE

C = controlled variable
D = drift input
G = transfer function of forward path
H = transfer function of feedback path
R = reference input signal

By examining the magnetic amplifier in block diagram form, the influence of the variations of the zero point may be examined for either magnetic feedback, Fig. 1, or electric feedback,[2] Fig. 2. That is, when the comparison between the reference and the feedback signal is made by opposing windings on the magnetic core of the amplifier, the zero shift may be introduced as shown in Fig. 1. On the other hand, if the reference and feedback are compared in an electric circuit external to the amplifier, Fig. 2 applies. In this figure, the dashed lines indicate that the drift introduction takes place within the amplifier components, i.e., internal to the amplifier block. In either case, the loop equations yield

$$\frac{C}{R+D}=\frac{G}{1+GH}$$

Thus it may be seen that both cases are equivalent to Fig. 3, where the loop regulates to a point erroneously offset from the reference position by the zero shift amount.

Accordingly, it may be concluded that offset of the zero or changes in the quiescent point introduced a fundamental error in the accuracy of the closed-loop system. This error is somewhat analogous to the resolution errors frequently encountered in mechanical systems. Since these errors are not reduced by increases in loop gain, the errors must be considered as a portion of the system design or application task, regardless of the cause producing the shift. Unfortunately, the nonlinear manner in which the various changes may combine makes it difficult to determine the total shift from information on shifts produced by separate causes.

For these reasons, it would be most helpful to those applying, using, or purchasing magnetic amplifiers to have the supplier indicate the total, i.e., maximum, change or shift in operating point arising from all causes, known or unknown, and especially including, but not restricted to, those changes arising from the anticipated variations of environment. From the foregoing analysis, the importance of distinguishing between quiescent shifts and gain variations becomes apparent. In keeping with current definitions, let me further suggest that it might be well to use the word drift to indicate this maximum quiescent change arising from all causes. Thus, limiting the use of the term "drift" to zero shifts, by excluding gain variations and by including all causes of quiescent change within the one term, should result in a more practical and useful definition.

REFERENCES

1. PROGRESS REPORT OF THE STANDARDS SUBCOMMITTEE OF THE AIEE MAGNETIC AMPLIFIER COMMITTEE, AIEE Committee Report. *AIEE CP 57-98* (copies available on request).
2. MAGNETIC AMPLIFIERS, Herbert F. Storm (book). John Wiley & Sons, Inc., New York, N. Y., 1955.

A Frequency-Converting Telephone Carrier Repeater for Military Use

G. GOLTSOS
NONMEMBER AIEE

J. H. JOHNSTON
NONMEMBER AIEE

R. B. ANDERSON
NONMEMBER AIEE

EARLY IN WORLD WAR II, the *CF* series of equipment was developed from commercial telephone carrier equipment for military tactical use. This equipment included the *CF-1* (telephone terminal) and *CF-3* (telephone repeater). The system was designed for 4-wire operation over the spiral-4 type of cable or radio relay.

Later in World War II, the need arose for converter-repeater equipment to enable the 4-channel *CF* system to operate over the many indigent 2-wire open-wire circuits. Converter *CF-4* and repeater *CF-5* were developed for this purpose.

Since that time, a series of compact, lightweight equipments has been designed specifically to meet the logistic and performance requirements of modern warfare. These equipments include the telephone terminal *AN/TCC-3* and telephone repeater *AN/TCC-5* (a 4-channel carrier system) and the telephone terminal *AN/TCC-7* and telephone repeaters *AN/TCC-8* and *AN/TCC-11* (a 12-channel carrier system).

The design criteria applied to all the equipment of this series were based on the proven technical art existing in the latter phase of the Korean period: aluminum transit cases, shock frames, plug-in chassis components, and vacuum-tube circuitry. Once the basic 4-channel system was designed, the converter-repeater *AN/TCC-6*, described in this paper, followed. Fig. 1 shows the *AN/TCC-6*, which was developed in 1952–1956, and which is more advanced in technical design. This, added to its important role in systems, qualifies the subject for technical release to interested engineering groups.

The equipment has been designed in accordance with the logistic requirement that fewer equipments perform the necessary system functions and that they be smaller and lighter. Converter-repeater *AN/TCC-6* replaces the system function of World War II converter *CF-4* and repeater *CF-5*. It accomplishes this replacement with an accompanying reduction in volume and weight so great that the *AN/TCC-6* might very well be continued as a companion piece with transistorized equipment now in laboratory phase of development. The converter-repeater *AN/TCC-6* weighs 70 pounds and has a volume of 2 cubic feet, as compared to the converter *CF-4* weight of 280 pounds in a volume of 12 cubic feet and the repeater *CF-5* weight of 380 pounds in a volume of 17.7 cubic feet.

Applications

The basic design function of the *AN/TCC-6* is to convert the 4-wire operation

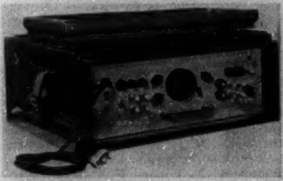


Fig. 1. The converter-repeater AN/TCC-6

Paper 58-215, recommended by the AIEE Wire Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted January 3, 1956; made available for printing November 18, 1957.

G. GOLTSOS and J. H. JOHNSTON are with the Signal Corps Engineering Laboratory, Fort Monmouth, N. J., and R. B. ANDERSON is with the Lenkurt Electric Company, San Carlos, Calif.

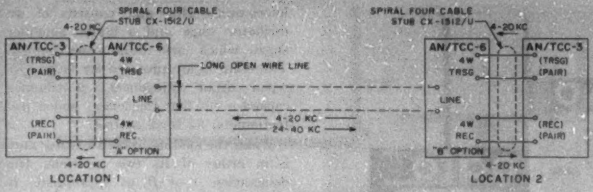

SPIRAL FOUR CABLE
STUB CX-1512/U
4-20 KC
AN/TCC-3
(TRSG)
(PAIR)
(REC)
(PAIR)
AN/TCC-6
4W
TRSG
LINE
4W
REC
"A" OPTION
LONG OPEN WIRE LINE
4-20 KC
24-40 KC
4-20 KC
LOCATION 1
SPIRAL FOUR CABLE
STUB CX-1512/U
4-20 KC
AN/TCC-6
4W
TRSG
LINE
4W
REC
"B" OPTION
AN/TCC-3
(TRSG)
(PAIR)
(REC)
(PAIR)
4-20 KC
LOCATION 2

Fig. 2. Block diagram of AN/TCC-6 used as 4-channel terminal

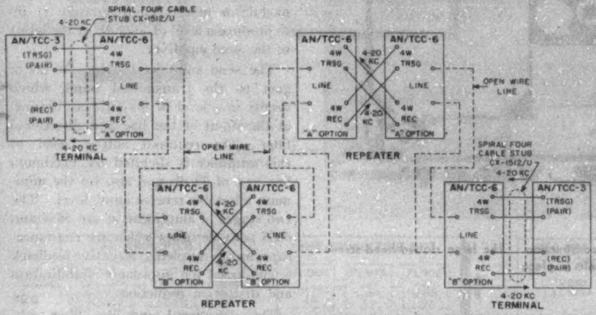

SPIRAL FOUR CABLE
STUB CX-1512/U
4-20 KC
AN/TCC-3
(TRSG)
(PAIR)
(REC)
(PAIR)
AN/TCC-6
4W
TRSG
LINE
4W
REC
"A" OPTION
4-20 KC
TERMINAL
OPEN WIRE
LINE
AN/TCC-6
AN/TCC-6
4-20 KC
4-20 KC
"A" OPTION
"A" OPTION
REPEATER
OPEN WIRE
LINE
"B" OPTION
"B" OPTION
REPEATER
SPIRAL FOUR
CABLE STUB
CX-1512/U
4-20 KC
AN/TCC-6
AN/TCC-3
(TRSG)
(PAIR)
(REC)
(PAIR)
"B" OPTION
4-20 KC
TERMINAL

Fig. 3. Block diagram of AN/TCC-6 used as 2-wire repeater

of telephone terminal *AN/TCC-3* and repeater *AN/TCC-5* to 2-wire operation. This is accomplished by translating the 4- to 20-kc base band for one direction of transmission to a 24- to 40-kc band. Thus, at the 2-wire side of the *AN/TCC-6,* the directions of transmission are separated in frequency rather than by two separate wire pairs. When 0.165 copper wire is used, a nominal nonrepeatered circuit length of 310 miles may be obtained. A block diagram of such an installation is shown in Fig. 2.

Two *AN/TCC-6* units may also be connected back to back, to form a 2-wire "frequency-frogging" repeater. By the process of frequency-frogging, the frequency bands for both directions of transmission are transposed and interchanged as they pass from one line section to the next. Fig. 3 illustrates this use in a multirepeatered system.

The technique of frequency-frogging offers two distinct advantages from the standpoint of transmission. As a result of the interchange of frequency bands, individual channels in the band are inverted as to their order in the frequency spectrum. Since attenuation is greater at the higher frequencies, signals subject to more loss in one section are subject to less loss in the next. The overall effect from terminal to terminal is a correction of the slope which would otherwise be present in the frequency-attenuation characteristic curve and, consequently, a uniform loss throughout both bands. In addition, singing and crosstalk between repeaters at the same installation are minimized, since the high-level output of each repeater is at a different frequency from that of the low-level input to itself and other repeaters.

For optimum performance, the number of repeater sections in a system is limited by design objectives to three. However, where some degradation is tolerable, longer systems may be used. Performance within design objectives may be obtained over distances up to 750 miles in a system using three repeater sections and 0.165 copper wire.

The *AN/TCC-6* may also be used with telephone terminal *AN/TCC-3* and radio set *AN/TRC-24* to form an 8-channel radio communication terminal. For this purpose, an *AN/TCC-6* is connected to each of two *AN/TCC-3* terminals and the two *AN/TCC-6* units are then cross-connected to each other.

Fig. 4 illustrates this application. The outgoing 4-channel, 4- to 20-kc band from terminal 2 is translated to a 24- to 40-kc band by one *AN/TCC-6* and then passed to the other *AN/TCC-6* where it is combined with the outgoing 4-channel, 4- to 20-kc band from terminal 1. The resulting 8-channel, 4- to 40-kc band is passed to the *AN/TRC-24* and transmitted. In a similar manner, the incoming 4- to 40-kc band from the *AN/TRC-24* is separated in one *AN/TCC-6* into a 4-channel, 4- to 20-kc band and a 4-channel, 24- to 40-kc band. The 4- to 20-kc band is directed to terminal 2 and the 24- to 40-kc band is passed to the other *AN/TCC-6* where it is translated to a 4- to 20-kc band and directed to terminal 1. Necessary cross connections are made with cords and jacks provided with the *AN/TCC-6.*

To facilitate trouble location and repair, all subassemblies of the *AN/TCC-6* are plug-in units; see Fig. 5. For ease of operation, switches and

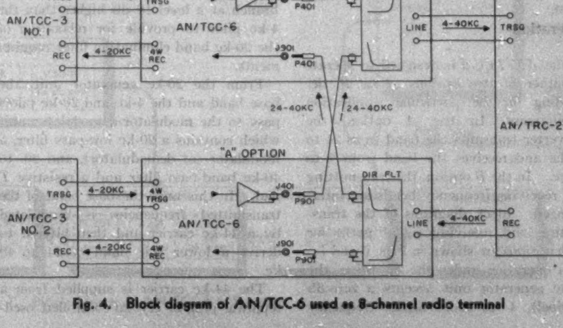

"B" OPTION
AN/TCC-3
NO. 1
TRSG
REC
4-20KC
4-20KC
4W
TRSG
4W
REC
AN/TCC-6
J401
P401
J901
P401
DIR FLT
LINE
4-40KC
TRSG
24-40KC
24-40KC
AN/TRC-24
"A" OPTION
AN/TCC-3
NO. 2
TRSG
REC
4-20KC
4-20KC
4W
TRSG
4W
REC
AN/TCC-6
J401
P901
J901
P901
DIR FLT
LINE
4-40KC
REC

Fig. 4. Block diagram of AN/TCC-6 used as 8-channel radio terminal

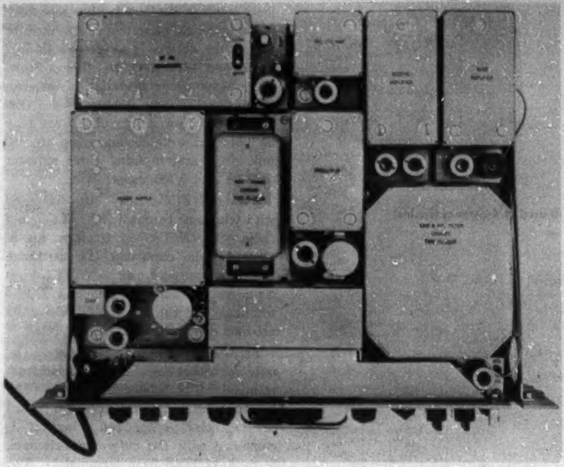


Fig. 5. Top view of AN/TCC-6 showing plug-in construction. The large slotted-head screws keep the plug-in units in place

controls have been kept to a minimum. Miniature electrical components were carefully chosen to meet the severe requirements of high- and low-temperature operation, vibration, shock, and moisture tests.

Consistent with the compact and portable nature of the new requirements for military carrier equipment, the *AN/TCC-6* weighs only 70 pounds and has over-all dimensions of 12 by 17 by 21 inches. The unit is mounted in a waterproof transit case for installation in locations where weatherproof shelter is not available. All controls, lamps, fuses, meters, etc., necessary for equipment lineup, maintenance, and operation are located on the front panel for convenient accessibility.

Operation

The *AN/TCC-6* is designed to operate in either of two options, *A* or *B*, depending on the particular application requirement. In the *A* option, the converter transmits the band from 24 to 40 kc and receives the band from 4 to 20 kc. In the *B* option, the transmitting and receiving frequency bands are interchanged. Block diagrams of the transmitting and receiving signal paths for both options are shown in Figs. 6 and 7.

In operation under the *A* option, the 20-kc generator unit accepts a zero-db (decibel), 4- to 20-kc carrier frequency band and a −25-db, 4-kc alarm pilot from an *AN/TCC-3* terminal. The purpose of the 20-kc generator unit is to provide attenuation control in the base band of 4 to 20 kc and to generate a 20-kc slope pilot from the 4-kc alarm and flat-gain pilot. Controls are provided in the unit for factory adjustment of the pilot level. A terminal/repeater switch is mounted in the top of the unit. When this switch is in the terminal position, the 20-kc pilot is generated for transmission through the system. When the switch is in the repeater position, the local 20-kc generator is disconnected and the 20-kc band elimination filter, *FL101*, is bypassed. Midband output level of this unit is −21 db. The 20-kc pilot is bridged onto the transmitting branch at a level 10 db higher than the 4-kc pilot to provide for relaxation of the 20-kc band elimination filter requirements.

From the 20-kc generator unit, the base band and the 4-kc and 20-kc pilots pass to the modulator-demodulator unit which contains a 20-kc low-pass filter, a modulator (or demodulator), and 20- to 40-kc band-pass filter and a resistive *T* pad. In this unit, the base band of the transmitted frequencies is modulated by a 44-kc carrier and then filtered to derive a lower side band of 24 to 40 kc.

The 44-kc carrier is supplied from a separate plug-in crystal-controlled oscillator unit. This unit consists of an oscillator stage and a buffer amplifier stage which are transformer-coupled. Two controls are provided for factory or repair depot adjustment; a trimmer capacitor for line frequency, and a potentiometer for level control.

When the modulator-demodulator unit is in either of its two positions, the designation *A* or *B*, as appropriate, is readable and a front panel lamp is lighted to indicate the option in use. The total loss of the modulator-demodulator unit is approximately 11 db to produce a level of −32 db at the input to the send amplifier.

The send amplifier provides 49 db of gain to the transmitted band, which results in a level of +17 db per channel at the input of the line and directional filter. The required output level of this amplifier is dictated by maximum line loss of 52 db and also by the minimum allowable receive input level. The two stages of amplification are *5654* and *6005* pentode tubes which are resistance-capacitance coupled. Negative feedback is employed for impedance stabilization and distortion reduction.

The directional and line filter subassembly gives directional separation of transmit and receive carrier frequencies, provides separation of the carrier and order wire circuits, and furnishes a balanced output for connection to the 2-wire line. In *A* operation a 24-kc high-pass filter is connected between the transmitting branch and the 2-wire line. Changing operation to the *B* mode (reversing the modulator-demodulator unit in its socket) changes the connections so that a 20-kc low-pass filter is connected between the transmitting branch and the line. Carrier and order wire circuits are separated at the line by means of a 4-kc high-pass filter and a 3-kc low-pass filter. The output of the 4-kc filter is passed through a transformer which provides a balanced circuit to the line. The 3-kc filter is a balanced configuration which flanks the transformer output. Order wire extension from the *AN/TCC-6* to the terminal equipment is provided by means of auxiliary equipment.

In the receiving direction, carrier frequencies from the line are passed through the directional filter to the receiving gain-adjusting unit, through coaxial connectors. Flat-gain controls consisting of a coarse-gain 20-db-per-step attenuator and a fine-gain 0- to 20-db variable attenuator, provide adjustment for different line conditions. At this point, a preamplifier of 10-db gain

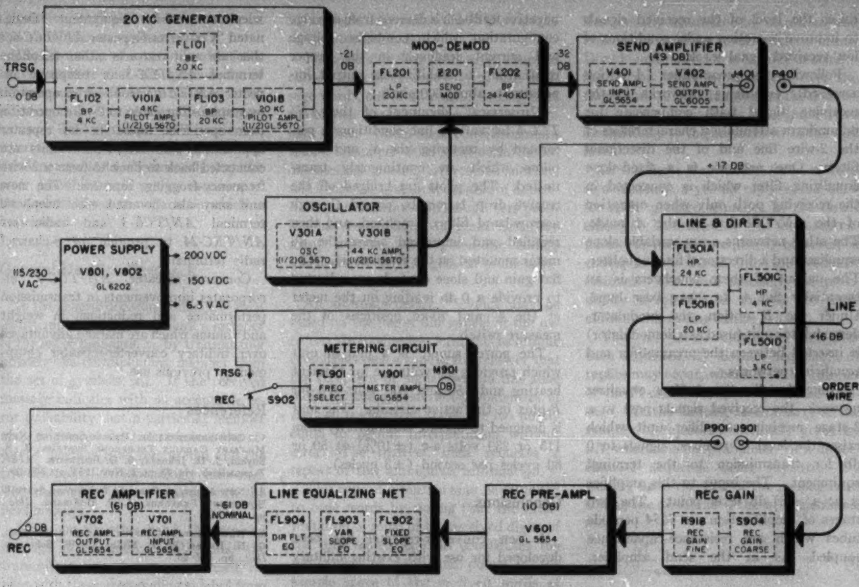

20 KC GENERATOR
FL101
BE
20 KC
TRSG
0 DB
FL102
BP
4 KC
V101A
4 KC
PILOT AMPL
(1/2) GL5670
FL103
BP
20 KC
V101B
20 KC
PILOT AMPL
(1/2)GL5670
-21 DB
MOD-DEMOD
FL201
LP
20 KC
Z201
SEND
MOD
FL202
BP
24-40 KC
-32 DB
SEND AMPLIFIER
(49 DB)
V401
SEND AMPL
INPUT
GL5654
V402
SEND AMPL
OUTPUT
GL6005
J401
P401
+17 DB
OSCILLATOR
V301A
OSC
(1/2)GL5670
V301B
44 KC AMPL
(1/2)GL5670
POWER SUPPLY
V801, V802
GL6202
115/230
VAC
200 VDC
150 VDC
6.3 VAC
LINE & DIR FLT
FL501A
HP
24 KC
FL501C
HP
4 KC
FL501B
LP
20 KC
FL501D
LP
3 KC
LINE
+16 DB
ORDER
WIRE
METERING CIRCUIT
FL901
FREQ
SELECT
V901
METER AMPL
GL5654
M901
DB
TRSG
REC
S902
P901
J901
REC AMPLIFIER
(61 DB)
V702
REC AMPL
OUTPUT
GL5654
V701
REC AMPL
INPUT
GL5654
-61 DB
NOMINAL
0 DB
REC
LINE EQUALIZING NET
FL904
DIR FLT
EQ
FL903
VAR
SLOPE
EQ
FL902
FIXED
SLOPE
EQ
REC PRE-AMPL
(10 DB)
V601
GL5654
REC GAIN
R918
REC
GAIN
FINE
S904
REC
GAIN
COARSE

Fig. 6. Block diagram of AN/TCC-6 in A option

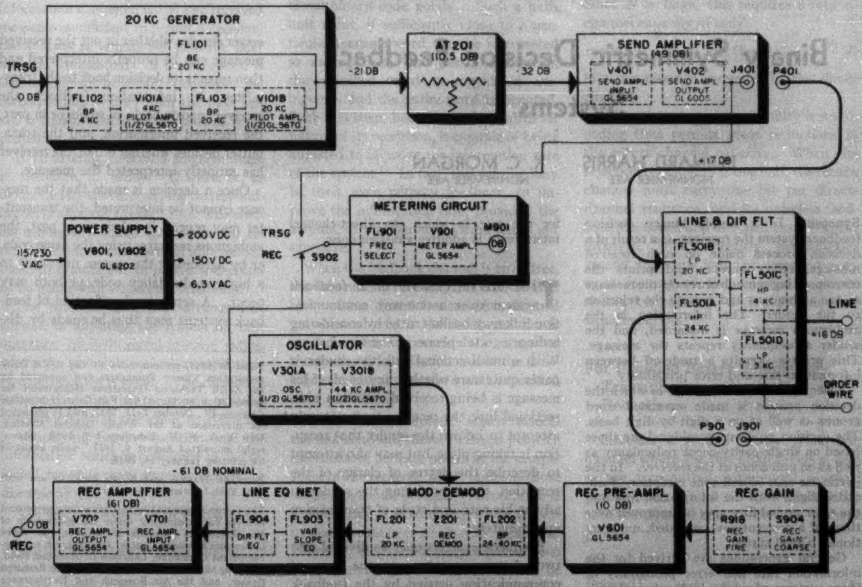

20 KC GENERATOR
FL101
BE
20 KC
TRSG
0 DB
FL102
BP
4 KC
V101A
4 KC
PILOT AMPL
(1/2)GL5670
FL103
BP
20 KC
V101B
20 KC
PILOT AMPL
(1/2)GL5670
-21 DB
AT201
(10.5 DB)
-32 DB
SEND AMPLIFIER
(49 DB)
V401
SEND AMPL
INPUT
GL5654
V402
SEND AMPL
OUTPUT
GL6005
J401
P401
+17 DB
METERING CIRCUIT
FL901
FREQ
SELECT
V901
METER AMPL
GL5654
M901
DB
TRSG
REC
S902
POWER SUPPLY
V801, V802
GL6202
115/230
VAC
200 VDC
150 VDC
6.3 VAC
LINE & DIR FLT
FL501B
LP
20 KC
FL501C
HP
4 KC
FL501A
HP
24 KC
FL501D
LP
3 KC
LINE
+16 DB
ORDER
WIRE
OSCILLATOR
V301A
OSC
(1/2) GL5670
V301B
44 KC AMPL
(1/2)GL5670
P901
J901
-61 DB NOMINAL
REC AMPLIFIER
(61 DB)
V702
REC AMPL
OUTPUT
GL5654
V701
REC AMPL
INPUT
GL5654
0 DB
REC
LINE EQ NET
FL904
DIR FLT
EQ
FL903
VAR
SLOPE
EQ
MOD-DEMOD
FL201
LP
20 KC
Z201
REC
DEMOD
FL202
BP
24-40 KC
REC PRE-AMPL
(10 DB)
V601
GL5654
REC GAIN
R918
REC
GAIN
FINE
S904
REC
GAIN
COARSE

Fig. 7. Block diagram of AN/TCC-6 in B option

raises the level of the received signals to improve operation under conditions of low received signal levels.

Following the preamplifier (*A* option assumed), equalizing networks in the receiving signal path compensate for nonuniform attenuation characteristics of the 2-wire line and of the directional filters. One network is a fixed-slope equalizing filter which is connected in the receiving path only when operation of the *AN/TCC-6* is in the *A* mode. The other networks are a variable slope equalizer and a directional filter equalizer. The output of these equalizers is an essentially flat 4- to 20-kc base band. Under the *B* option, the modulator-demodulator unit (used as a demodulator) is inserted between the preamplifier and equalizing networks.

From the directional filter equalizer network, the received signals pass to a 2-stage receiving amplifier unit which raises the level of received signals to 0 db for transmission to the terminal equipment. The input to this amplifier is at a −61-db level point. The two stages of amplification are *5654* pentode tubes which are resistance-capacitance coupled. As in the send amplifier, negative feedback is derived from a bridge configuration which combines voltage and current feedback in the proper relationship to stabilize the output impedance to about 600 ohms.

"In-service" alignment of the *AN/TCC-6* for various line conditions is performed by metering the 4- and 20-kc pilots which are continuously transmitted. The pilots are bridged off the receive drop terminals, passed through narrow-band filters, amplified, and then rectified and impressed across the db meter mounted on the front panel. The flat gain and slope controls are adjusted to provide a 0-db reading on the meter at the 4- and 20-kc positions of the measure switch.

The power supply is a plug-in unit which provides 6.3 volts a-c for filament heating and 150 and 200 volts d-c for *B*-plus in the active circuits. The unit is designed to operate satisfactorily from 115 or 230 volts a-c (±10%) at 50 or 60 cycles per second (±3 cycles).

Conclusions

A new converter-repeater has been developed for use with existing military telephone carrier equipment. Designated "Converter-Repeater *AN/TCC-6*," this new unit converts either telephone terminal *AN/TCC-3* or telephone repeater *AN/TCC-5* from 4-wire operation over spiral-four cable to 2-wire operation over open-wire lines. In its repeater application, two *AN/TCC-6* units are connected back to back to form a 2-wire frequency-frogging repeater. The new unit may also be used with telephone terminal *AN/TCC-3* and radio set *AN/TRC-24* to provide an 8-channel radio terminal.

Converter-repeater *AN/TCC-6* incorporates improvements in transmission performance and reductions in weight and volume which are marked advantages over military converter-repeater equipment in previous use.

References

1. Considerations for Development of New Military Carrier Telephone Systems, R. S. Boykin, J. H. Johnson, S. D. Bedrosian. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 503–08.

2. New Military Carrier Telephone Systems Equipment Features, J. P. Hoffmann. *Ibid.*, pp. 509–15.

3. New Military Carrier Telephone Systems, G. H. Huber, W. F. Miller, C. W. Schramm. *Ibid.*, pp. 515–25.

Binary Symmetric Decision Feedback Systems

BERNARD HARRIS
NONMEMBER AIEE

K. C. MORGAN
NONMEMBER AIEE

Synopsis: In a simple binary decision feedback system the receiver, as a result of a decision process made on each received message, either accepts and prints the corresponding symbol, or rejects the message as too ambiguous and reports the rejection to the sender. The information in the ambiguous message is discarded, and the sender subsequently repeats the message. This process permits a trade-off between information rate and error probability.

Schemes are considered here in which the decision process is made on coded word groups as well as on a digit-by-digit basis. The decision processes considered are those based on single-parity-check redundancy as well as on null zones at the receiver. In the particular case of null-zone reception of uncoded digits both the information rate and the error probability may be improved over those of a conventional, uncoded, unidirectional, bilevel system.

General expressions are derived for the information rate and error probability of these systems when the feedback process is error-free. Specific examples are calculated for the case in which the direct-channel interference is additive gaussian noise.

THE ADVANTAGE of a feedback system over a one-way communication link may be illustrated by considering ordinary telephone communication. With a unidirectional link the sender is never quite sure whether any or all of his message is being received. With a bidirectional link the receiver may not only attempt to inform the sender that reception is taking place, but may also attempt to describe the degree of clarity of the reception, thus permitting the sender to adjust his rate and style of transmission appropriately.

Feedback systems may be divided into two classes, according to the type of communication carried by the feedback channel. In decision feedback the receiver decides whether or not the received message can be properly interpreted and then reports its decision back to the transmitter. In information feedback the receiver reports back, in whole or in part, the received information, and the transmitter decides whether or not the receiver has properly interpreted the message.

Once a decision is made that the message cannot be interpreted, the transmitter may repeat, in whole or in part, the ambiguous message, using the same code, or by agreement the system may shift to a higher redundancy code, or both may occur. A separate classification of feedback systems may thus be made by dis-

---

Paper 57-1041, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957, and re-presented for discussion at the Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted August 9, 1957; made available for printing January 14, 1958.

Bernard Harris and K. C. Morgan are with New York University, New York, N. Y.

The research reported in this paper was performed at New York University under the sponsorship of the Electronics Research Directorate, Air Force Cambridge Research Center, Air Research and Development Command, Laurence G. Hanscom Field, Bedford, Mass., the Air Force Cambridge Research Center, and the Air Research and Development Command, Baltimore, Md., under Contract *AF 19(604)-1049*.

tinguishing between fixed-code systems and variable-code systems.

The advantage of a variable-code feedback system is that the redundancy of the code may be efficiently matched to the interference. This permits the setting of an upper limit on the probability of error, whereas for a unidirectional link the redundancy must be chosen such that the error probability will not be excessive at the poorest signal-to-noise ratio which is anticipated.

General Decision Feedback

In the general decision feedback system an attempt is made to transmit a member of the set of messages, x_i, by transmitting initially the corresponding member from the set of symbols, x_{1i}. If the received message indicates with an acceptable error probability that a particular member of x_i originated the transmission, the received message is so interpreted and the transmitter is correspondingly informed. The transmitter then proceeds with the next message.

When, on the other hand, the received message is considered ambiguous, the transmitter, upon being so informed, replies with the corresponding symbol from the set x_{2i}. In a similar manner, if the information contained in the two received messages is sufficient to permit a decision, the process is terminated and the next message transmitted. If the reception is still insufficient to permit a decision, the corresponding symbol from the next set, x_{3i}, is transmitted, and the process is repeated.

In an attempt to send a single message x_i, therefore, the sender will transmit, as demanded by the receiver, the corresponding set of symbols $x_{1i}, x_{2i}, \ldots$. The process will be terminated only when the receiver indicates that the total information received is sufficient to permit a decision to be made with a specified reliability.

The simplest fixed-code arrangement consists of adjusting the system for best operation at high signal-to-noise ratios, where the probability of requiring corrective information is small. The corrective information may then be poorly encoded without excessive reduction in efficiency by employing simple repetition, and the information which is contained in the rejected message may correspondingly be discarded.

The decision to accept or reject a message is, in general, based on some scheme for distinguishing those points in sample space closely associated with specific message points from those points which are too ambiguous. Such a scheme, for example, may utilize one of the various error-detecting codes.

The general coding problem is, then, to choose, for a given code length, such message words, out of all possible words, that the chosen words are at least $2k+1$ from each other. Under these conditions more than k errors per word must occur before one message is misinterpreted as another, and hence the system is capable of detecting up to k errors. Although this problem has been studied in great detail by many authors, and systematic coding schemes evolved, the general problem remains unsolved.

An alternate decision scheme may be based on the use of null zones[1] at the detector. In its simplest form this technique establishes such a zone at the receiver that the reception of signals within this zone indicates the a posteriori probabilities of the two transmitted states are too nearly equal to permit interpretation without error. The receiver then requests a repeat of this ambiguous information.

This type of reception may be extended to code words by noting that the interpretation of some of the received digits of the code words as nulls constitutes the recognition of additional points in the sample space of the received words at half-unit distances from the conventional binary code points. Such a half-unit point, if sufficiently close to a particular message word, may be interpreted as arising from that word. Otherwise, the received word is considered ambiguous, and the entire word is repeated. Such a system, because of the multilevel nature of its reception, is capable of being adjusted to improve the information rate of the system.[1] In fact, this scheme may be used, even without feedback, to improve the information rate, provided the amount of ambiguous information is not excessive.

When the feedback channel is error-free, the information rate and the error probability of the direct transmission channel remain unaltered if the feedback loop is opened. The important advantage of the feedback loop is that the entire message is eventually received at the specified error probability.

Double-null reception, which bisects the null zone and thereby retains polarity information, permits points in the sample space of the received words between those of the single-null system. This scheme is thus capable of further improving the information rate over single-null reception. However, single-null reception has been shown, under certain conditions, to be capable of achieving about one half of the improvement in information rate theoretically attainable by increasing the number of receiver levels without limit. Double-null reception is capable of achieving only a small improvement over the single null. One advantage of the double-null scheme, however, is the comparative insensitivity to the adjustment of the null level.

In the most elementary form of decision feedback a yes-no reply is fed back for each message sequence in the direct channel. Considerably less feedback capacity is actually required, however, if coding is introduced into the feedback channel. The more precisely the probability of rejection is known, the more efficient this coding can be.

Thus, if a yes-no reply is fed back after each sequence, 2^N distinct sequences of replies may occur, where N is the number of direct-channel sequences in the message. In the limiting case of large N the number of rejections is closely given by uN, where u is the probability of a direct-channel sequence being rejected. By waiting until the end of the message and coding the positions of the uN rejections, only

$$\frac{N!}{(uN)!(N-uN)!} \tag{1}$$

distinct possible replies are required. Since N is large, this requires a return-channel capacity of only

$$-u\log_2 u-(1-u)\log_2(1-u) \tag{2}$$

bits per message sequence in the direct channel.

When the rejection probability is small, coding thus permits great reductions in the return-channel capacity. When the rejection probability is one half, the return channel must carry one bit per direct-channel sequence, and for noiseless feedback the coded and uncoded cases become identical. For rejection probabilities greater than one half the accepted positions instead of the rejected positions should be encoded, and a symmetrical situation exists.

For Uncoded Binary Symmetric Transmission: Single-Null Reception

Consider a general binary symmetric channel where the two input levels x_1 and x_2 are transmitted with equal a priori probabilities. With a 3-level decoder the received signal is assigned to one of three groups:

$$\begin{array}{ll} \text{Group I} & y \geq h \\ \text{Group II} & -h < y < h \\ \text{Group III} & y \leq -h \end{array} \tag{3}$$

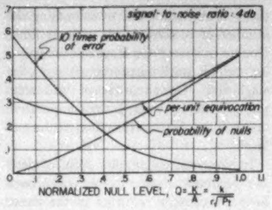


Fig. 1. Variation of parameters of decision feedback system using single-null reception of uncoded binary pulses

Group I records as x_1, Group III records as x_2, and group II records as O.

Because of the symmetry of the system, and presuming amplitude symmetry of the noise, the information rate is bits per symbol is

$$R(k)=p(\mathrm{I}/x_1)\log_2\frac{2p(\mathrm{I}/x_1)}{p(\mathrm{I}/x_1)+p(\mathrm{III}/x_1)}+ p(\mathrm{III}/x_1)\log_2\frac{2p(\mathrm{III}/x_1)}{p(\mathrm{I}/x_1)+p(\mathrm{III}/x_1)} \quad (4)$$

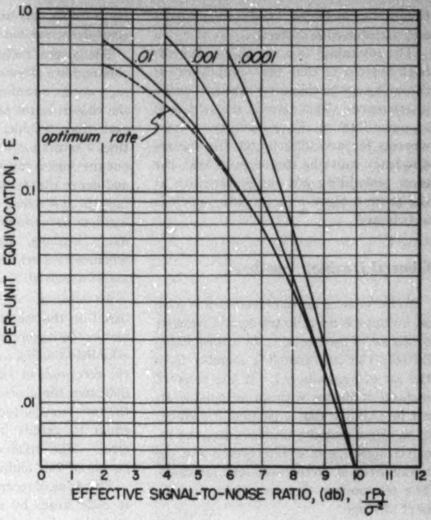
This information rate will in general be greater than that attainable by simple bilevel reception. With a unidirectional system this increase in rate is achieved at the cost of receiving parts of the message as ambiguous. With a feedback link the ambiguous portions of the message may be repeated as often as necessary until properly received. In the particular case of an error-free feedback channel this improvement in information rate may be maintained while the entire message is received at some specified reliability. The error probability of the received message is then

$$P_e=\frac{p(\mathrm{I}/x_1)}{p(\mathrm{I}/x_1)+p(\mathrm{III}/x_1)} \quad (5)$$

Subject only to the reasonable constraint that $p(\mathrm{I}/x_1)$ exceed $p(\mathrm{III}/x_1)$ it may be shown that, when the noise is additive-symmetrical, the rate attains its maximum at some $k>0$. This is shown by noting that dR/dk is positive at $k=0$. In a similar manner it may be readily shown that P_e also attains its best value at some $k>0$. Hence it is always possible to find some setting of the null level at which both the rate and error probability are improved.

To determine the degree of improvement requires that a specific type of interference be considered. For simplicity the interference will again be considered to be additive gaussian noise. A binary symmetric channel is assumed in which there are r transmitted samples per bit of information. When the interference is additive gaussian and is uncorrelated at the sampling points, each received r-group with the same summed amplitude has the same a posteriori probability, and thus can be identified with a single effective received level z.

Fig. 2. Per-unit equivocation versus signal-to-noise ratio for uncoded binary decision feedback system with single null adjusted for a given probability of error. Direct-channel interference is additive gaussian noise. Feedback channel error-free



When the null-zone boundaries are $\pm k$, a normalized signal amplitude A and a normalized boundary K may be conveniently defined as

$$A=\left(\frac{rP_t}{2\sigma^2}\right)1/2$$

$$\text{and } K=\frac{k}{\sqrt{(2\sigma^2 r)}}$$

where P_t is the signal power and σ^2 the noise power.

$$p(\mathrm{I}/x_1)=1/2[1+\phi(A-K)] \quad (6)$$

$$p(\mathrm{III}/x_1)=1/2[1-\phi(A+K)]$$

where $\phi(x)$ is the error function

$$\phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}dx$$

The analytic expression for the per-message-symbol rate for single-null reception may then be found from relation 4.

Fig. 1 shows the corresponding manner in which the per-unit equivocation and the error probability of the accepted message vary as the null level is raised. The subsequent improvements are, of course, the result of the presence of the feedback channel, which fills in the rejected portions of the message. The degree of activity of the feedback channel is shown by the probability-of-nulls relation in Fig. 1.

Fig. 2 shows the extent by which the rate will vary when the feedback system is adjusted for operation at a given error probability. It should be noted that, for a sufficiently high signal-to-noise ratio and a fixed error probability, the reception reduces to that of a single-null system at the optimum rate. This optimum rate is, of course, better than that of unidirectional bilevel transmission.

For example, at a signal-to-noise ratio of 6 db (decibels), the unidirectional bilevel system has an information rate of 0.84 and an error probability of 0.06. The feedback system, adjusted for optimum rate, has an information rate which is about 4% better and an error probability which is 0.01. By the introduction of a sufficiently high null level the error probability may be reduced tenfold to 0.001 with a corresponding rate reduction of about 10%. To reduce the error probability a hundredfold to 0.0001 would require a rate reduction of 40%.

To examine the nature of this exchange between error probability and information rate, consider the system at low signal-to-noise ratios. The information rate then approaches

$$R \doteq \frac{2}{\pi \ln 2}\left[\frac{\epsilon^{-2K^2}}{1-\phi(K)}\right]^2 \tag{7}$$

and the optimum rate adjustment becomes

$$R \doteq 1.17A^2 \tag{8}$$

Unfortunately the corresponding error probability is excessively high, at 0.50.

If the null level is adjusted for a low probability of error instead of for the optimum rate, the rate relation simplifies to

$$R \doteq \frac{2KA^2}{\sqrt{\pi}\ln 2}\epsilon^{-K^2} \tag{9}$$

and the error probability to

$$P_e \doteq \frac{\epsilon^{-(A+K)^2}}{2(A+K)\sqrt{\pi}} \tag{10}$$

whence R and P_e are related by

$$R \doteq \frac{2A^2P_e}{\ln 2}\left(\frac{-\ln P_e}{\pi}\right)^{1/2} \tag{11}$$

This relation shows that, at low signal-to-noise ratios and at low error probabilities, the information rate is roughly proportional to the error probability.

At high signal-to-noise ratios the rate reduction is considerably less. Unfortunately the relation between error probability and per-unit equivocation is not as simple.

Thus

$$E \doteq \frac{\epsilon^{-(A-K)^2}}{2(A-K)\sqrt{\pi}} \tag{12}$$

while

$$P_e \doteq \frac{\epsilon^{-(A+K)^2}}{2(A+K)\sqrt{\pi}} \tag{13}$$

These relations show, as might be expected, that at high signal-to-noise ratios multifold improvement in the error probability may be obtained with only a small change in the null level. While the per-unit equivocation will correspondingly show multifold increase, the information rate will remain approximately constant as long as the absolute value of the per-unit equivocation is low. This phenomenon is characteristic of most coding schemes, namely, that the schemes operate best when the signal-to-noise ratio is reasonably high.

Hamming Single-Error-Detecting Decision Feedback

SINGLE-NULL RECEPTION

Binary transmissions are usually concerned not with individual pulses, but rather with groups of pulses representing the symbols of a particular alphabet. It is therefore important to perform an analysis based on n-digit pulse groups or words, where the per-unit equivocation and the error probability are based on whether or not the entire words as entities are properly interpreted and recorded. It should be noted that for languages having interword constraints the analysis would necessarily be based on sequences of words as entities.

Consider a Hamming[2] code of m message digits, plus a single additional digit for parity check, being used in a decision feedback system employing single-null reception. The feedback channel is, as before, error-free. Normally the presence of the extra digit permits the detection of an odd number of errors.

The registration of single nulls permits the single parity check to be used, not only for single-error detecting, but substantially for single-error correcting as well. Thus, if a single digit of the $(m+1)$ digits is registered as a null, the parity check is used to fill in the null, and the resultant word is printed. When two or more nulls are received, the word is not printed. When none of the digits are nulls, the word is printed if and only if parity checks.

The symbols q, p, and u are defined as the probabilities that the word is printed correctly, printed incorrectly, or not printed respectively. These are related by

$$q+p+u=1 \tag{14}$$

Similarly, q_0, p_0, and u_0 represent the corresponding single-digit probabilities of being correct, incorrect, or receiving a null. In the presence of additive gaussian noise having a flat spectrum these become

$$\begin{aligned} q_0 &= \frac{1}{2}[1+\phi(A-K)] \\ p_0 &= \frac{1}{2}[1-\phi(A+K)] \\ u_0 &= \frac{1}{2}[\phi(A+K)-\phi(A-K)] \end{aligned} \tag{15}$$

Assume that all permutations of the m message digits are included in the group of words or alphabet which may be sent. Let Q_r represent the probability

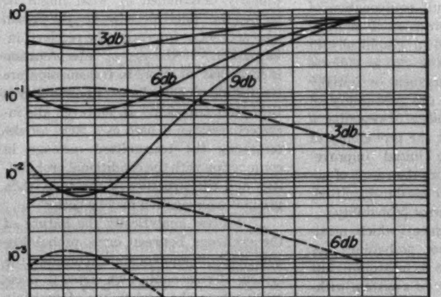


Fig. 3 (left). Per-unit equivocation and probability of error of accepted message for various signal-to-noise ratios and null levels for 5-message-digit, single-parity-check, binary decision feedback system using single-null reception. Direct channel interference is additive gaussian noise. Feedback channel error-free.

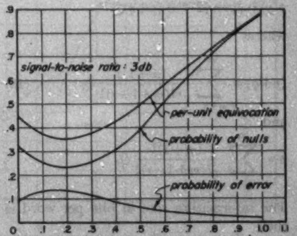


Fig. 4 (right). Variations of parameters of decision feedback system using single-null reception and 5-message-digit, single-parity-check, binary transmission as null level is varied

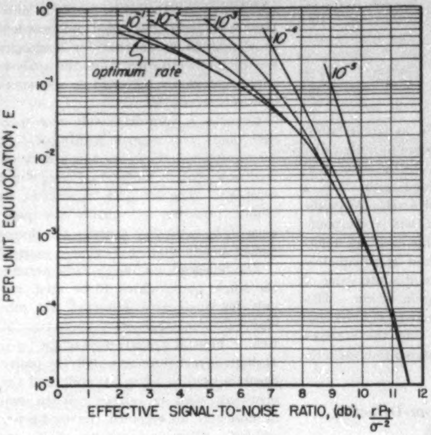


Fig. 5. Per-unit equivocation versus signal-to-noise ratio for a 5-message-digit, plus single-parity-check binary decision feedback system with single null adjusted at values above optimum for a given probability of error. Direct-channel interference is additive gaussian noise. Feedback channel error-free

that a transmitted word will be accepted as another word r units away. Since this requires the misinterpretation of an even number of digits, r is even. If no nulls are received, this misinterpretation can occur only when r errors are made. A misinterpretation may also occur, when one null is received, if r additional errors are made plus a correct interpretation of the null or if $(r-1)$ errors are made, plus an incorrect interpretation of the null. Therefore

$$Q_r = q_0^{m+1-r} p_0^r + (m+1-r) q_0^{m-r} u_0 p_0^r + r q_0^{m+1-r} u_0 p_0^{r-1} \qquad (16)$$

The corresponding number of words r units away from the transmitted word is

$$\frac{(m+1)!}{r!(m+1-r)!} \qquad (17)$$

When the input messages are equally probable and the noise is symmetrical, the output messages are also equally probable, and

$$p(x) = 2^{-m}$$

$$p(y) = (1-u)2^{-m} = (p+q)2^{-m}$$

where

$$(p+q) = \sum_{r=0,2,\ldots}^{m} \frac{(m+1)!}{r!(m+1-r)!} Q_r \qquad (18)$$

and

$$p = \sum_{r=2,4,\ldots}^{m} \frac{(m+1)!}{r!(m+1-r)!} Q_r$$

The information rate in bits per word is then

$$R = m(p+q) - (p+q) \log_2 (p+q) + \qquad (19)$$

$$\sum_{r=0,2,\ldots}^{m} \frac{(m+1)!}{r!(m+1-r)!} Qr \log_2 Qr$$

while the corresponding error probability is given by

$$p_e = \frac{p}{(p+q)} \qquad (20)$$

The relation between the error probability and the per-unit equivocation of this system, in the presence of additive gaussian noise, is shown graphically in Figs. 3, 4, and 5. The curves indicate that as the null adjustment is initially increased from zero, the probability of nulls, i.e., the probability of rejecting a word, decreases, while the probability of error increases. The initial improvement in rate is therefore the result of accepting more words per word transmitted, albeit with increased error probability.

At the null adjustment which maximizes the rate the error probability has already passed through its maximum, but is still poorer than that of a system without nulls. When the null level is adjusted to approximately twice its optimum value, the per-unit equivocation of the system with nulls about equals that of the system without nulls. The corresponding probabilities of nulls are also about equal, but the error probability is now smaller. This means that when the null level is adjusted for the same rate as the system without nulls, about the same number of words are accepted per repeat, but the error probability of the accepted words is smaller.

By adjustment of the null level for a value above the optimum rate adjustment it is possible to exchange information rate for error probability over wide limits. Below the optimum rate adjustment such an exchange is also possible, but not on as efficient a basis.

Consider the system at a signal-to-noise ratio of about 6 db. Without nulls the per-unit equivocation is about 0.1, while the error probability is 0.005. By introducing null reception and maximizing the rate, the rate is improved by about 5% while the error probability is increased 20%. However, if the null system is adjusted for the same rate as the system without nulls, the error probability is decreased by 20%. Adjusting the null level to values above the optimum rate adjustment can result in a fivefold reduction in error probability, i.e., to 0.001, at the cost of 70% rate reduction.

It is significant to note that here is an example of a system which, on the basis of per-unit equivocation, is better, at the optimum rate adjustment, than conventional binary reception, but which, on the basis of error probability, is worse. The reason is that error probability and per-unit equivocation are not uniquely related to each other. Specifically, the error probability is the same whether only one incorrect word or many different words can be received per transmitted word. The relative value of per-unit equivocation and error probability is ultimately determined by what might be termed the "final" decoder, i.e., the person who acts on or interprets the received information. If the act or interpretation is such that all errors in transmission are treated identically, the error probability is the better criterion. If, however, the incorrect message is used as a basis for determining the transmitted message in conjunction with the conditional probabilities and the structure of the code, then per-unit equivocation may be better.

To express analytically the nature of the exchange between error probability and per-unit equivocation, consider the system at reasonably high signal-to-noise ratios where the probability of multiple errors or multiple null digits per word is small. It may be shown that the information rate approaches

$$R \doteq mQ_0 = mq \qquad (21)$$

while the error probability becomes

$$P_e \doteq \frac{m(m+1)Q_2}{2} \qquad (22)$$

From relation 21 it may be established that operating the system at the optimum rate adjustment causes an improvement in the signal-to-noise ratio equivalent to

1.4 db when compared with the system without null reception on the basis of rate. As previously shown, however, the error probability at this optimum rate adjustment is poorer. In fact, at high signal-to-noise ratios, and at the optimum rate adjustment

$$-\ln P_e \doteq 3 \ln u_0$$

or

$$-\ln P_e \doteq 2.06A^2 \tag{23}$$

This is in contrast to simple binary reception where, at high signal-to-noise ratios

$$-\ln P_e = 2A^2 \tag{24}$$

The corresponding deterioration in the error probability is equivalent to a loss of 0.26 db in power, whereas the increase in rate is equivalent to a 1.4-db increase in power.

The behavior of the null adjustment which maximizes the error probability may be determined by examining the system at high signal-to-noise ratios, where

$$P_e \doteq \frac{m(m+1)Q_2}{2} \tag{25}$$

Since

$$Q_2 \doteq p_0^2 + 2u_0p_0 \tag{26}$$

the consequent relation is

$$-\frac{u_0'}{p_0'} = 1 + \frac{u_0}{p_0} \tag{27}$$

For additive gaussian noise this reduces to

$$\exp(4AK) \doteq \frac{\phi(A+K) - \phi(A-K)}{1 - \phi(A+K)} \tag{28}$$

or in the limit

$$A^2 \doteq -\frac{\ln Q}{4Q} \tag{29}$$

This latter relation shows that, as the signal-to-noise ratio increases, the null level at which the error probability achieves its maximum decreases to zero. This is in contrast with the adjustment for maximum rate, wherein the normalized null level Q approaches 17.2% in the limit. The comparison confirms the observation that at a given signal-to-noise ratio, as the null level is raised, the error probability achieves its maximum before the rate is maximized. Hence, best operation is obtained for null adjustments greater than that required to optimize rate.

There remains to examine the condition of adjustment at which the system with nulls has the same rate as conventional reception. It is noted that, at high signal-to-noise ratios, adjustments of the null above the optimum rate adjustment correspond to u_0^2 being greater than p_0. Since, with null reception

$$E = (m=+1)p_0 + \frac{m(m+1)}{2}u_0^2 + \dots \tag{30}$$

then, at high signal-to-noise ratios, above the optimum rate adjustment

$$E \doteq \frac{m(m+1)}{2}u_0^2 \tag{31}$$

The condition for this adjustment may consequently be shown to reduce to

$$A^2 \doteq 2(A-K)^2 \tag{32}$$

or

$$Q \doteq 1 - \frac{\sqrt{2}}{2} = 29.3\% \tag{33}$$

By adjusting the null level to 29.3% of the effective signal-to-noise ratio, in the presence of additive gaussian noise, the rate of the system with nulls will be nearly equal to the rate of a system employing conventional reception.

It is of interest to compare the probability of rejecting a word for these two systems. Simply by noting that, for this adjustment

$$(1-Q_0) >> Q_2 >> Q_4 \dots \tag{34}$$

it follows that, at high signal levels

$$u \doteq 1 - Q_0 \tag{35}$$

and also

$$E \doteq 1 - Q_0 \tag{36}$$

Thus, at high signal-to-noise ratios, the equivocation is primarily dependent on the rate of rejection rather than the distribution of error.

It is interesting to compare the relations for error probability of the feedback systems with and without null reception. Without nulls

$$-\ln P_e \doteq 2A^2 \tag{37}$$

whereas, with nulls

$$-\ln P_e \doteq (5 - 2\sqrt{2})A^2 \tag{38}$$

when the rates are adjusted to be equal. The introduction of null reception has improved the error probability by an equivalent 0.35-db increase in power. The nulls are thus performing error-correction.

The behavior of this system at low signal-to-noise ratios may be improved, as may all systems, by the introduction of a variable code. Consider, for example, that the system, at low signal-to-noise ratios, is to be modified so that each digit is repeated several times by the sender and correspondingly integrated before decoding. For noise which is uncorrelated with the signal, this results in an effective improvement in the signal-to-noise ratio, but causes a corresponding decrease in rate. The optimum arrangement, from the viewpoint of best rate, is shown in Fig. 6. For example, below a signal-to-noise ratio of about 3 db, an error probability of 0.01 can be maintained at a higher information rate by a system which employs single repetition. Below about 0.5 db, double repetition is better. At still lower signal-to-noise ratios additional repetitions are desirable.

One of the main advantages of the feedback system over a unidirectional system is that this altering of the redundancy of the code can be done as the interference level changes. Hence the higher redundancy required at low signal-to-noise ratios need not reduce the information rate at high levels.

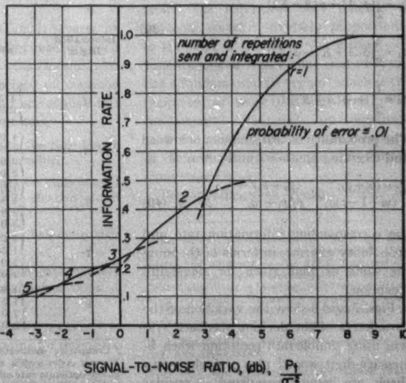


Fig. 6. Information rate versus signal-to-noise for a variable-code, 5-message-digit, single-parity-check, binary decision feedback system. Direct-channel interference is additive gaussian. Feedback channel error-free

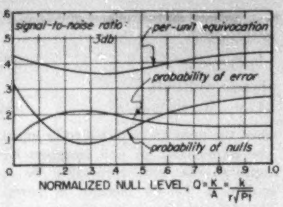


Fig. 7. Variations of parameters of decision feedback system using double-null reception and 5-message-digit, single-parity-check, binary transmission as null level is varied

DOUBLE-NULL RECEPTION

With double-null reception the interpretation of the received signal is identical to that of single-null reception except when multiple nulls are being received. Then, instead of always being rejected, the word is rejected only if parity does not check the digits indicated by the polarities of the nulls. With such a decoder the probability of error cannot be continuously reduced by increasing the null level, since very high null levels reduce the system to simple binary reception. One advantage of this system, however, is that the information rate for any setting of the null adjustment is better than that of a system without nulls.

As before, let q_0 and p_0 represent the probabilities of receiving correct and incorrect digits respectively, while q_1 and p_1 represent the probabilities of receiving the corresponding nulls. In the presence of additive gaussian noise

$$
\begin{aligned}
q_0&=\frac{1}{2}[1+\phi(A-K)]\\
q_1&=\frac{1}{2}[\phi(A)-\phi(A-K)]\\
p_1&=\frac{1}{2}[\phi(A+K)-\phi(A)]\\
p_0&=\frac{1}{2}[1-\phi(A+K)]
\end{aligned}
\tag{39}
$$

The probability of transmitting one word and receiving another r units away is

$$
\begin{aligned}
Q_r=&(q_0+q_1)^{m+1-r}(p_0+p_1)^r+\\
&(m+1-r)q_0^{m-r}p_0^rp_1+rq_0^{m+1-r}p_0^{r-1}q_1
\end{aligned}
\tag{40}
$$

the corresponding information rate and probability of error, in terms of Q_r, being the same as that given for single-null reception.

Figs. 7 and 8 show the variation of the parameters of the decision feedback system using double-null reception when 5-message-digit words, plus a single parity check, are being transmitted. Consider Fig. 7 where the variation is shown for a 3-db additive gaussian noise signal-to-noise ratio. The effect of introducing the null level is to reduce the probability of nulls, but to increase the probability of error. This means that more words are accepted without being repeated, but the probability of error of the accepted words is greater. However, for a fixed signal-to-noise ratio, the range of variation of both the per-unit equivocation and the probability of error is limited. This is in contrast with uncoded single-null reception where, by the introduction of the null, both the per-unit equivocation and the probability of error are improved, and it is also in contrast with a single-parity-check, coded, single-null reception system where, by varying the null level, the rate may be indefinitely exchanged for probability of error.

To express analytically the nature of this exchange, consider the system at a high signal-to-noise ratio where the probability of multiple error digits or multiple nulls per word are small. Then

$$Q_0>>Q_2>>Q_4\ldots \tag{41}$$

The information rate, as for single-null reception, then approaches

$$R\doteq mQ_0 \tag{42}$$

while the probability of error of the accepted digits becomes

$$P_e\doteq\frac{m(m+1)}{2}Q_2 \tag{43}$$

Thus for double-null reception, as for single-null, the procedure of maximizing the rate may be approximated by maximizing the Q_0 or q.[1] Differentiating the expression for q yields, at high signal-to-noise ratios

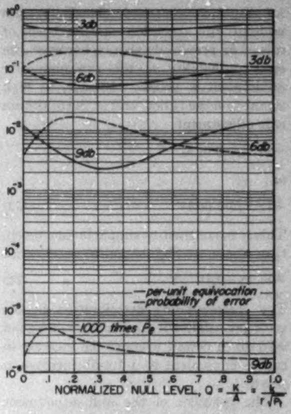


Fig. 8. Per-unit equivocation and probability of error at accepted message for various signal-to-noise ratios and null levels for a 5-message-digit, single-parity-check, binary decision feedback system using double-null reception. Direct-channel interference is additive gaussian noise. Feedback channel error-free

$$\frac{q_0'}{p_1'}=\frac{-q_0}{mp_1} \tag{44}$$

or, in the presence of additive gaussian noise

$$\exp(4AK)\doteq\frac{1+\phi(A-K)}{m[\phi(A+K)-\phi(A)]} \tag{45}$$

This relation for the optimum null

Table I. Comparison of Error Probability*

Information Digits	Signal-to-Noise Ratio A†	Signal-to-Noise Ratio Db	Uncoded	Hamming Code	Wagner Code	Decision Feedback
4	1.0	3.01	0.209	0.191	0.143	0.0816
	1.5	6.53	0.0349	0.0248	0.0115	0.00420
	2.0	9.03	0.00313	0.00145	0.00030	0.00008
	3.0	25.5	42×10^{-7}	6×10^{-7}	$<10^{-7}$	1.3×10^{-9}
5	1.0	3.01	0.269	0.310	0.190	0.115
	1.5	6.53	0.0493	0.0513	0.0164	0.00617
	2.0	9.03	0.00486	0.00375	0.00045	0.00012
	3.0	25.5	84×10^{-7}	25×10^{-7}	$<10^{-7}$	2.0×10^{-9}
6	1.0	3.01	0.325	0.335	0.236	0.150
	1.5	6.53	0.0641	0.0530	0.0220	0.00860
	2.0	9.03	0.00673	0.00346	0.00062	0.00017
	3.0	25.5	138×10^{-7}	17×10^{-7}	$<10^{-7}$	2.8×10^{-9}
7	1.0	3.01	0.377	0.362	0.282	0.192
	1.5	6.53	0.0789	0.0552	0.0281	0.0112
	2.0	9.03	0.00871	0.00330	0.00082	0.00022
	3.0	25.5	201×10^{-7}	12×10^{-7}	$<10^{-7}$	3.9×10^{-9}
8	1.0	3.01	0.425	0.388	0.326	0.231
	1.5	6.53	0.0937	0.0580	0.0347	0.0142
	2.0	9.03	0.01075	0.00322	0.00103	0.00029
	3.0	25.5	272×10^{-7}	9×10^{-7}	$<10^{-7}$	5.1×10^{-9}

* Comparing undirectional systems employing no coding, Hamming single-error-correcting codes, or Wagner codes with a decision feedback system using a single-parity-check code and single-null reception at the optimum rate adjustment.

† This is identical to the α in Table III of reference 3.

level is not the same as that for the single-null system. In particular, at high signal-to-noise ratios, the expression reduces to

$$e^{4AK} = \frac{2A\sqrt{\pi}}{m} e^{A^2} \tag{46}$$

For large A, since the exponents predominate, this reduces to

$$Q = \frac{K}{A} = \frac{1}{4} \tag{47}$$

To understand better the effect of this choice of null level, note that the optimum adjustment approximately corresponds to adjusting p_0 to be equal to the product of q_1p_1. Based on this adjustment the per-unit equivocation becomes

$$E \doteq (m+1)(p_0 + mq_1p_1) \tag{48}$$

or

$$-\ln E \doteq \frac{25}{16} A^2$$

Based on per-unit equivocation, the introduction of the double null can, therefore, yield an equivalent improvement of 1.9 db over the system without nulls, and a 0.5-db improvement over single-null reception. The corresponding probability of error of the accepted message at the maximum rate adjustment is, however, somewhat poorer than for the system without nulls. The difference in error probabilities, decibelwise, vanishes at large signal-to-noise ratios.

Compared to Unidirectional Systems Which Use Error-Correcting Codes

It is of interest to compare the performance of the single-null decision feedback system with unidirectional systems which use error-correcting codes. The schemes chosen for comparison are those tabulated by Silverman and Balser.[3] These are: 1. an uncoded m-digit word transmission; 2. a single-error-correcting code using the minimum number of parity checks; and 3. an m-digit Wagner code. The feedback code chosen for comparison is the Hamming single-parity-check code of m information digits. The null level will be considered to be adjusted for optimum rate.

Table I reproduces the foregoing information, adding the data for the feedback system. These systems are compared on the basis of the same noise-free rates. It should be noted that, because of the rejection feature, the actual time required to transmit a message using the decision feedback system is greater than that required by the undirectional systems by a factor equal to the reciprocal of $(1-u)$. This corresponds, over the range of m in this table, to an increase of from 23% to 40% for $A=1$ and from 4% to 8% for $A=1.5$. For greater signal-to-noise ratios the time devoted to repetition becomes negligible. The tabulation emphasizes the reduction in error probability which may be achieved by using decision feedback for error correction.

It is seen therefore that, where a return channel is already available, or where the need for reliability justifies the additional cost of establishing such a channel, the decision-feedback approach offers a relatively simple and practicable means of achieving a substantial error reduction.

References

1. IMPROVEMENT OF BINARY TRANSMISSION BY NULL-ZONE RECEPTION, **F. J. Bloom, S. S. L. Chang, B. Harris, A. Hauptschein, K. C. Morgan.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, July 1957, pp. 963–75.

2. ERROR DETECTING AND ERROR CORRECTING CODES, **R. W. Hamming.** *Monograph no. 1757*, Bell Telephone System, New York, N. Y., Apr. 1950.

3. CODING FOR CONSTANT-DATA-RATE SYSTEMS, PART I, A NEW ERROR-CORRECTING CODE, **R. A. Silverman, M. Balser.** *Proceedings*, Institute of Radio Engineers, vol. 42, Sept. 1954, pp. 1428–35.

A Constant-Amplitude Random Function Generator

G. A. HELLWARTH
NONMEMBER AIEE

THE constant-amplitude random function generator described in this report resulted from the need for an electronic generator which would combine some of the properties of random noise waveshapes with the known characteristics of periodic, constant-amplitude waves. In this instance, a wave shape was desired which had as a main feature a constant peak-to-peak amplitude, but whose path between peaks was in some way randomly chosen.

At first inspection the simplest way to generate such a function would be that of amplitude clipping any noise-like wave shape. However, this was thought to be unsatisfactory because the wave spends too much time at the peak value, or clipping level, rather than traversing between peaks. The next criterion on the wave shape, then, was that it reverse its direction immediately after reaching each peak.

There appeared to be several common wave shapes that would accomplish this task. The most obvious were:

1. A sawtooth wave with the time length of its tooth randomly determined.
2. A triangular wave with the time lengths of both the positive and negative slopes randomly chosen.
3. A cosine wave with its half-cycle time length randomly chosen.
4. A wave with a random path between peaks, each path maintaining a slope of constant sign.

At this point, it becomes necessary to define some standard terms for discussion of the proposed wave shape.

Given a constant-amplitude wave shape, a single excursion of the wave between two peaks shall be called a "side." The time derivative or slope of a side is a constant for a sawtooth or triangle wave. A side of a cosine wave shall be the value of the cosine function between 0 and π, or π and 2π. The side time-length, or "side time," is defined as the time duration of a side's excursion from one peak to the next. The side time of a triangle or sawtooth wave is inversely proportional to the slope of the side. A constant side time triangle or cosine wave (not random) has a fundamental frequency which is the reciprocal of twice the side time. The fundamental frequency or repetition rate of a sawtooth wave is the reciprocal of the side time, assuming a zero flyback time; see Fig. 1.

Perhaps the most evident reason for

Paper 58-533, recommended by the AIEE Electronic and High Frequency Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958. Manuscript submitted January 27, 1958; made available for printing February 25, 1958.

G. A. HELLWARTH is with the University of Michigan, Ann Arbor, Mich.

This work was sponsored by the United States Army Signal Corps.

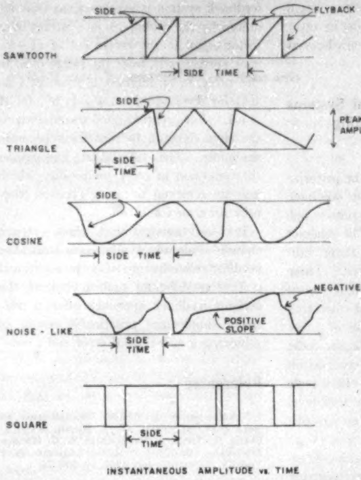


Fig. 1 (left). Constant - amplitude random functions

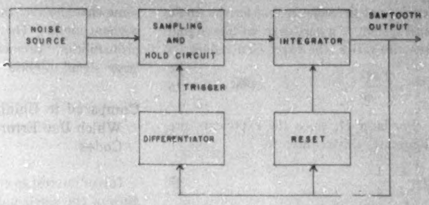


Fig. 2 (above). Block diagram of random sawtooth generator

the introduction of this terminology is that the functions will actually be generated by randomly choosing each side time-length, resulting in the description of the wave being a probability distribution of side times. In practice, however, it was found to be much easier to measure the power spectrum than the side-time distribution, so both frames of reference will necessarily be used.

The Random Sawtooth Generator

General Description of Operation

The first type of random wave shape generated was the sawtooth, primarily because this type of generator was thought to be the simplest to design and build. The generator was to operate at audio frequencies and was to have as much flexibility as possible in controlling its side time-length distribution. The random sawtooth block diagram is shown in Fig. 2, and a schematic circuit diagram is shown in Fig. 3. Essentially, the generator consists of a Gaussian noise source being sampled at the time of the sawtooth flyback, the noise sample being held and integrated until the output reaches some constant value. At this point the integrator is reset, causing the sawtooth flyback, and a new noise sample is taken and integrated, etc. The integrator is a constant-current pentode discharging a capacitor, resetting being due to a thyratron recharging the capacitor. The noise samples are capacitor-coupled into the pentode grid, a controllable grid bias determining the center side time; the center side time is that side time resulting from a zero noise sample; see equation 4.

Integrator Operation

A simplified integrator circuit, as shown in Fig. 4, is helpful in observing the generator operation during the formation of a single side.

The current sources in Fig. 4 represent a pentode whose only plate load is a capacitor C. Thus, during the length of time required to generate one side, a constant current is being fed into the capacitor to produce the ramp function of voltage across the capacitor. If g_m is the transconductance of the pentode, $I_o=g_mE_o$ is the pentode current resulting from the grid bias voltage E_o, and $I_n=g_mE_n$ is the pentode current resulting from the noise sample voltage E_n, then the capacitor voltage $E(t)$ will be

$$E(t)=\frac{I_o+I_n}{C}\,t=\frac{g_m(E_o+E_n)}{C}\,t,\quad 0\leq t\leq T \tag{1}$$

And since (E_o+E_n) is constant during a side, the time length of the side is that time T, required for the capacitor voltage to change by E_m volts, the peak-to-peak voltage of the sawtooth wave. Therefore

$$T=\frac{CE_m}{g_m(E_o+E_n)}=\frac{CE_m}{I_o+I_n} \tag{2}$$

Thus, equation 2 gives the relation between the sample of noise E_n, and the side time T, produced by such a noise sample. The effect of varying the pentode grid bias E_o can also be seen. A quantity to note in equation 2 is the transconductance g_m, which in practical cases is never a constant, but is a function of E_n. By choosing differing types of pentodes in the generator circuit, one is able to obtain different relations between E_n and T according to the particular way a pentode transconductance varies with its input. Another tool available for controlling the side time T, produced by a noise sample E_n, is the nonlinear amplitude attenuating filter, which can be inserted between the noise source and the sampling circuit. This type of network will effectively introduce to equation 2 a nonlinear factor, a monotonic function of E_o, which multiplies E_o.

The circuit described in Fig. 3 contains a sharp cut-off pentode and no amplitude attenuating network. Amplitude attenuating filters are discussed in greater detail in a later section.

Discussion of Laboratory Model

The random sawtooth circuit is relatively stable and reliable. The most critical element appears to be the pentode, since its d-c characteristics are used to determine the center side time. However, the output power spectrum is not seriously affected by slight changes in the pentode characteristics. A limit can be set on the maximum time length of any side by simply paralleling the integrating capacitor with a suitable resistance so that if the pentode were cut off by a large noise sample, the resistor would discharge the capacitor at a predetermined exponential decay rate.

The center side time is controllable between 0.2 and 10 milliseconds by varying the pentode grid bias. Maximum deviations of side time from the center depend on the pentode grid bias setting, but 200 to 1 maximum to minimum side-time ratios are achievable with the pentode grid bias near the middle of its range. A much more satisfactory way of changing the center side time would be to vary the size of the integrating capacitor, always

Vo
Pt.
19

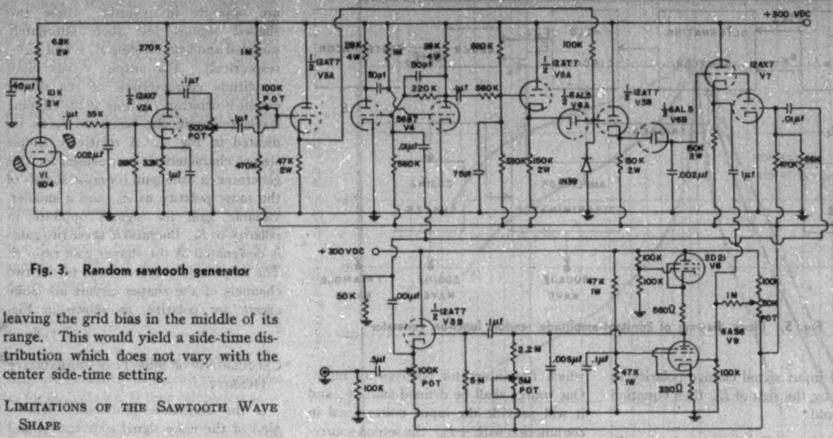

6.8K
2W
40μf
10K
2W
.1μf
39K
.002μf
V1
6D4
270K
½ 12AX7
V2A
.1μf
39K
3.3K
1μf
500K
POT
1M
100K
POT
470K
½ 12AT7
V3A
47K
2W
28K
4W
1M
50pf
50pf
28K
4W
220K
5687
V4
.01μf
560K
.1μf
560K
680K
75pf
330K
150K
2W
½ 12AT7
V5A
½ 6AL5
V6A
100K
1N39
½ 12AT7
V5B
½ 6AL5
V6B
150K
2W
.002μf
60K
2W
12AX7
V7
1μf
.01μf
470K
56K
+300 VDC
-150 VDC
+300VDC
50K
.001μf
½ 12AT7
V3B
.1μf
.5μf
100K
100K
POT
5M
2.2M
5M
POT
.005μf
.1μf
47K
1W
47K
1W
100K
100K
2D21
V8
560Ω
6AS6
V9
330Ω
100K
1M
100K
50K
POT

Fig. 3. Random sawtooth generator

leaving the grid bias in the middle of its range. This would yield a side-time distribution which does not vary with the center side-time setting.

LIMITATIONS OF THE SAWTOOTH WAVE SHAPE

The output power spectrum of the random sawtooth is very wide and contains a considerable band above audio frequencies. (There are no spectrum or side-time curves in Appendix I for the sawtooth, since it became only a forerunner of the more significant triangle and cosine generator.) The higher frequency spectral power, caused chiefly by the rapid flyback of the sawtooth, requires a correspondingly wide bandwidth for any associated circuit to pass the wave and maintain its constant-amplitude character. It was then decided to try to generate one or more of the other waves described in Fig. 1, waves which would not have the large amount of high-frequency power resulting from discontinuities. In the triangle and cosine waves, the power spectrum was expected to be much more concentrated in the range corresponding to the side-time distribution, resulting also in increased ability to control the output power spectrum.

The Random Triangle, Cosine, and Square Wave Generator

DESIGN CONSIDERATIONS

In planning the design of a random triangle generator, there were several problems of circuit choice and complexity that were not encountered in the sawtooth generator. The basic block diagram of the circuit, as shown in Fig. 5, is similar to the diagram of the sawtooth, but the contents of the blocks are quite different.

The first design consideration resulted in choosing an operational-amplifier integrator circuit which could integrate positive and negative inputs. The output of the integrator operates between constant peak-to-peak limits by the action of the amplitude discriminator circuit. When the integrator output reaches a certain magnitude, i.e., the preset peak value, the bistable discriminator circuit is "flipped," resulting in a trigger pulse to the sampler and a reversal of polarity of the integrator input signal. This action takes place at both positive and negative peaks, thus confining the integrator output between the peak values. The output to the integrator is the sum of two signals, (1) the constant-amplitude square-wave output of the amplitude discriminator E_o, and (2) the random-amplitude square-wave output of the noise sampling and holding circuit E_n. The operation of the circuit is wholly dependent upon the successful generation of the integrator input signal.

Analysis of Generator Operation

RELATION BETWEEN INTEGRATOR INPUT AND OUTPUT SIDE TIME

Given an integrator circuit as shown in Fig. 6(A), an equation of operation similar to equation 1 can be written. The time T required for E to change by E_m volts is given by equation 3, since (E_o+E_n) is constant throughout the interval T.

$$T=RC\frac{E_m}{E_o+E_n} \tag{3}$$

The "center" side time T_o is again defined as that side time generated when the noise voltage sample E_n is zero.

$$T_o\triangleq RC\frac{E_m}{E_o} \tag{4}$$

Fig. 6(B) shows the expected ramp function of voltage versus time obtained at the integrator output, when a constant is integrated. The differing times required for the output to change by E_m volts when the input is varied is indicated by the three sides pictured. Fig. 6(C) is a quasi-log-log plot of equation 3.

INTEGRATOR INPUT REQUIREMENTS

The first requirement to be noted is that at every proposed peak of the integrator output, the polarity of the integrator input signal (E_o+E_n) must be reversed. The signal E_o, which is a positive constant, is intended to produce the necessary changes of polarity at the integrator input by alternately appearing as $+E_o$ or $-E_o$. However, in order that

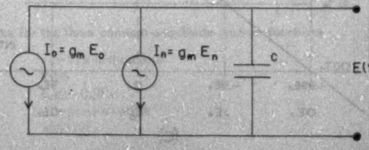

$I_o = g_m E_o$
$I_n = g_m E_n$
C
E(t)

Fig. 4. Simplified equivalent circuit for generating a side

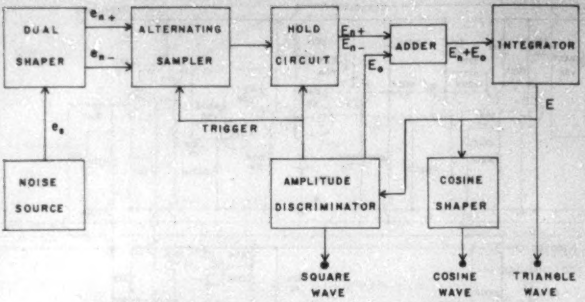

DUAL SHAPER
e_{n+}
e_{n-}
ALTERNATING SAMPLER
HOLD CIRCUIT
E_n+
E_n-
ADDER
E_o
E_n+E_o
INTEGRATOR
E
e_s
TRIGGER
NOISE SOURCE
AMPLITUDE DISCRIMINATOR
COSINE SHAPER
SQUARE WAVE
COSINE WAVE
TRIANGLE WAVE

Fig. 5. Block diagram of constant-amplitude random function generator

the total input signal change polarity as dictated by the sign of E_o, then equation 5 must hold:

$$E_n > -E_o \text{ when } +E_o \text{ is integrated} \quad (5)$$
$$E_n < E_o \text{ when } -E_o \text{ is integrated}$$

If side times shorter than $1/2(T_o)$ are to be desired, then there must be two separate sources of noise signal E_n from which the integrator receives its input. One source shall be defined as E_n+, and it will provide the input noise signal in conjunction with $+E_o$; the second source, E_n-, will provide the input with $-E_o$.

The dual noise shaper takes a single noise voltage input e_s, and feeds this signal through two parallel shaping amplifiers (or attenuators) whose characteristics are opposite in polarity. These two shaped signals are then alternately sampled and held yielding E_n+ and E_n-, respectively. The output versus input amplitude characteristic of the shaper circuit is arbitrary, as long as the clipping levels maintain the limits on e_n as indicated in Fig. 7. A relatively simple shaper characteristic was chosen for this generator: a fixed gain for noise signals of the same polarity as E_o, and a smaller, variable gain for signals opposite in polarity to E_o; the ratio of these two gains is designated as the shaper gain ratio R. The amplitude characteristics of the two channels of the shaper circuit are alike except for polarity, as shown in Fig. 7.

CALCULATION OF SIDE-TIME PROBABILITY DENSITY

If the amplitude probability density $p(e_s)$ of the noise signal e_s is known and the shaping circuit characteristic is chosen, the side-time probability density can be calculated; that is, the probability that the output wave is being generated with a side time within some finite range

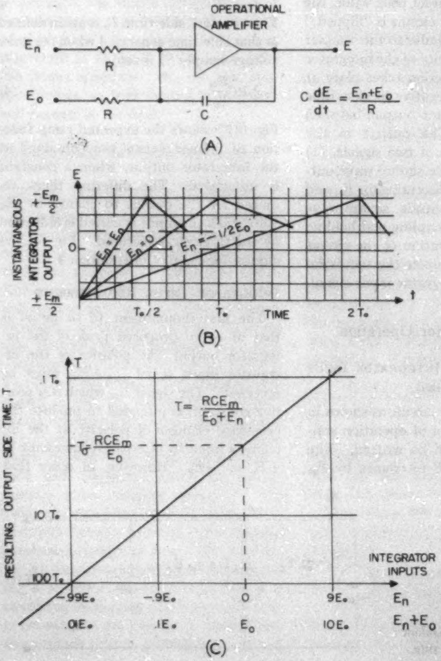

OPERATIONAL AMPLIFIER
E_n
R
E_o
R
C
E
$C\frac{dE}{dt}=\frac{E_n+E_o}{R}$
(A)
E
INSTANTANEOUS INTEGRATOR OUTPUT
$+\frac{-E_m}{2}$
0
$\pm\frac{E_m}{2}$
$E_n=E_o$
$E_n=0$
$E_n=-1/2E_o$
$T_o/2$
T_o
TIME
$2T_o$
t
(B)
T
RESULTING OUTPUT SIDE TIME, T
$1T_o$
$10T_o$
$100T_o$
$T=\frac{RCE_m}{E_o+E_n}$
$T_o=\frac{RCE_m}{E_o}$
INTEGRATOR INPUTS
$-.99E_o$
$-.9E_o$
0
$9E_o$
E_n
$.01E_o$
$.1E_o$
E_o
$10E_o$
E_n+E_o
(C)

Fig. 6 (left). Operation of integrating circuit

A—Integrator circuit
B—Generating side
C—Integrator characteristic

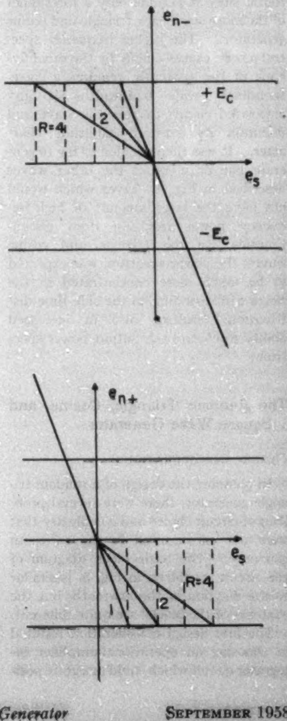


Fig. 7 (right). Dual noise shaper characteristics. Plot of output e_n versus input voltage e_s for different gain ratios R with a fixed clipping level E_c

of values. Since the shaping characteristic is a nonlinear function, as shown in Fig. 7, any calculation will have to be done piecewise with each linear portion of the shaper characteristic.

In general, the amplitude probability density of the shaped noise $p(e_n)$ is found by

$$p(e_n) = p(e_s)\frac{de_s}{de_n} \tag{6}$$

where the derivative is calculated from the shaping characteristic. From equation 4 the side time T, resulting from a sample E_n of the shaped noise voltage e_n, is given by

$$T = \frac{T_oE_o}{E_o + E_n} \tag{7}$$

The probability density of choices of side time $q(T)$ made by the circuit as it samples e_n for each choice, is then calculated as

$$q(T) = p(E_n)\frac{dE_n}{dT} \tag{8}$$

where $p(E_n)$ is assumed equal to $p(e_n)$ and the derivative dE_n/dT is found from equation 7:

$$\frac{dE_n}{dT} = \frac{d}{dT}\left(\frac{T_oE_o}{T} - E_o\right) = -\frac{T_oE_o}{T^2} \tag{9}$$

Since each side time chosen by the circuit occupies a portion of time equal to T, the output side-time probability density $r(T)$ is found by weighting the $q(T)$ density with the factor T:

$$\begin{aligned} r(T) &= \frac{T}{T_a}q(T) = \frac{T}{T_a}p(E_n)\frac{dE_n}{dT} \\ &= \frac{T}{T_a}p(e_s)\frac{de_s}{de_n}\frac{dE_n}{dT} \\ &= \frac{T_oE_o}{T_aT}p(e_s)\frac{de_s}{de_n} \end{aligned} \tag{10}$$

where T_a is a constant used to maintain unity area under the $r(T)$ curve. Thus equation 10 shows the method for calculating the side-time probability density of the output constant-amplitude random function.

EXPERIMENTAL DETERMINATION OF THE OUTPUT POWER SPECTRUM

It would be desirable to know the relationship between the side-time probability density, $r(T)$ (previously determined), and the output power spectrum. Much time was spent in an attempt to relate these two functions, but without success. However, in spite of the advantages of having the mathematical equation, some practical insight into this relationship can be derived experimentally. When the noise signal e_s is gaussian noise, the measured power spectrum for a cosine

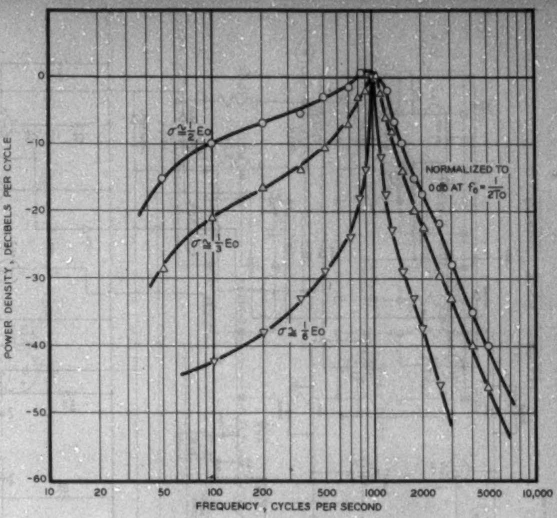


Fig. 8. Cosine wave power spectra as noise voltage e_s is varied

$R = 1$
$E_c \cong 0.9E_o$
$T_o = 1/2$ millisecond
e_s = gaussian noise

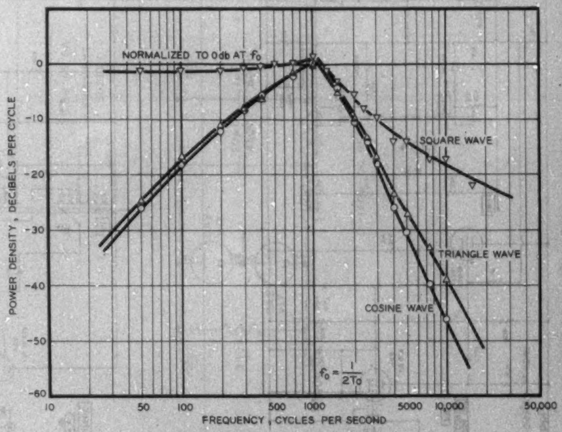


Fig. 9. Power spectra for the three constant-amplitude random functions

$T_o = 1/2$ millisecond
$R \cong 5$
$E_c \cong -0.9E_o$
e_s = gaussian noise
$\sigma \cong 1/2E_o$

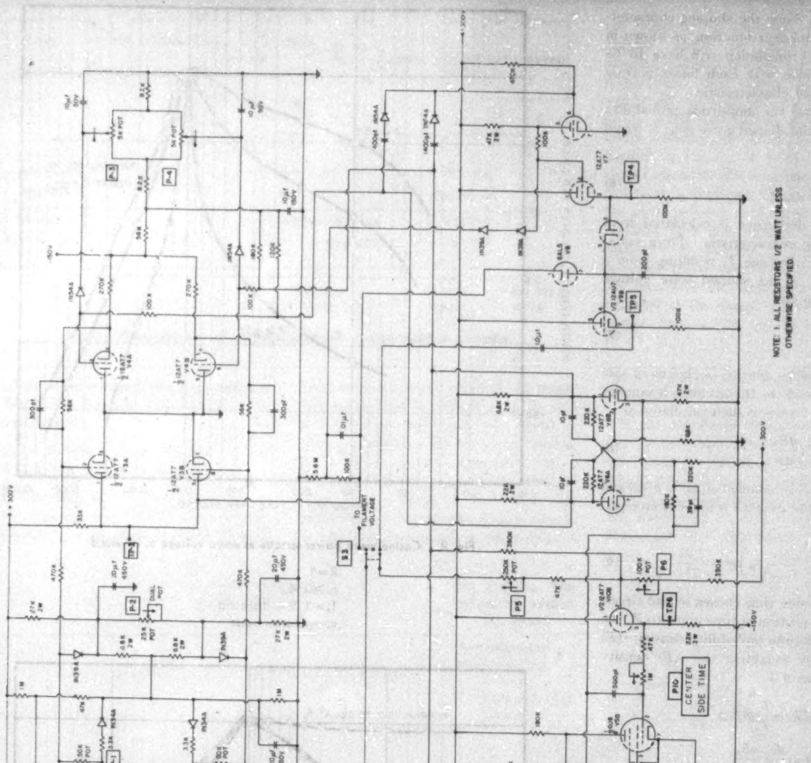

NOISE GAIN CONTROL
NOISE OUT
TO FILAMENT VOLTAGE
CENTER SIDE TIME
OUTPUT GAIN CONTROL
OUTPUT
+300V
-150V
NOTE: 1. ALL RESISTORS 1/2 WATT UNLESS OTHERWISE SPECIFIED.

Fig. 10. Constant-amplitude random function generator

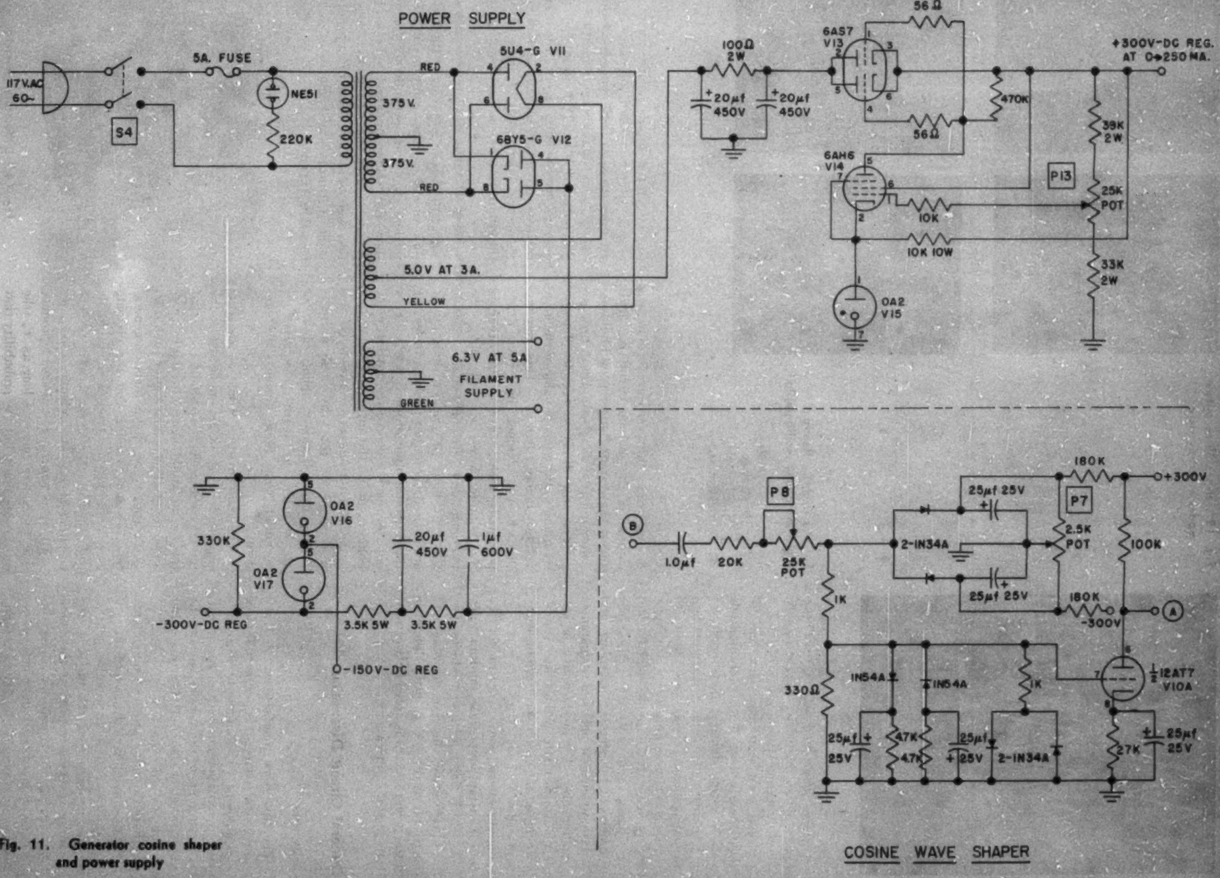

POWER SUPPLY
117V.AC 60~
S4
5A. FUSE
NE51
220K
375V.
375V.
RED
RED
5U4-G V11
6BY5-G V12
5.0V AT 3A.
YELLOW
6.3V AT 5A
FILAMENT SUPPLY
GREEN
100Ω 2W
20μf 450V
20μf 450V
6AS7 V13
56Ω
56Ω
470K
6AH6 V14
10K
10K 10W
OA2 V15
+300V-DC REG. AT 0→250 MA.
39K 2W
25K POT
P13
33K 2W
330K
OA2 V16
OA2 V17
20μf 450V
1μf 600V
3.5K 5W
3.5K 5W
-300V-DC REG
-150V-DC REG
B
1.0μf
20K
P8
25K POT
2-1N34A
25μf 25V
25μf 25V
P7
2.5K POT
180K
180K
100K
+300V
-300V
A
1K
330Ω
1N54A
1N54A
1K
25μf 25V
47K
47K
25μf 25V
2-1N34A
½ 12AT7 V10A
27K
25μf 25V
COSINE WAVE SHAPER

Fig. 11. Generator cosine shaper and power supply

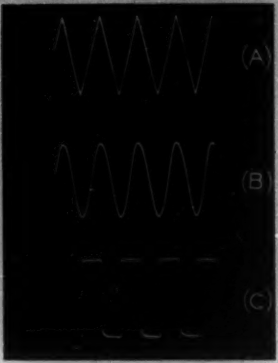


Fig. 12 (left). Three output modes at steady state

A—Triangle
B—Cosine
C—Square

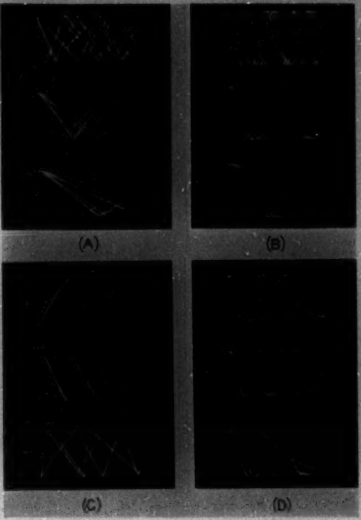


Fig. 13 (right). Random triangle and cosine oscillographs

A—Triangle
B—Cosine
C—Triangle
D—Cosine

output, as shown in Fig. 8, consists primarily of two relatively straight roll-offs from the center frequency $f_o = 1/2(T_o)$. It would then seem necessary only to relate these roll-offs with the amplitude of the side-time density on either side of T_o. Unfortunately, the constructed laboratory generator was not built to provide easy measurement of the noise rms voltage σ, the clipping level E_c, and the shaper gain ratio R. Also, the accuracies of the shaping, sampling, and hold circuits are not sufficient to guarantee a high degree of reproducibility of settings. Approximate values are stated, however, on the power spectrum measurements shown.

Fig. 9 shows power spectra for the three constant-amplitude random functions which the generator produces (i.e., the square wave, the triangle wave, and the cosine wave). Very little difference in the power spectra is noted between the triangle and cosine waves.

General Circuit Discussion

COMMENTS

The laboratory generator was designed and constructed as an experimental piece of equipment. Consequently, there are many potentiometers built in for adjusting parameters in the various parts of the circuit, most of the controls being either for gain or for d-c bias.

Perhaps the major problem encountered in the design and construction of the constant-amplitude random function generator was insuring that the input to the integrator reversed polarity at the end of each side. In the circuit developed, this was done, as mentioned previously, in the noise-shaping circuit by the clipping diodes. Since the noise wave form in each channel of the shaping circuit is unbalanced, the shaping, sampling and hold circuitry must be direct-coupled in order to preserve the direct voltage level corresponding to the average value of the noise voltage before shaping. This requires that the two channels hold their direct potentials, with respect to each other, or the clipping levels E_c will vary; the tolerances on E_c are fairly small if the side times as large or greater than 10 T_o are desired. Another problem in feeding the generator input properly is whether or not to capacitor-couple the output of the hold-circuit to the integrator input. A large capacitor was used in the circuit, but it is suspected that the small currents flowing through this coupling capacitor are sufficient to disturb the capacitor potential, since these currents are random.

A recommendation would be that the clipping level E_c be accomplished at the integrator input (see Appendix II) so that the circuit could more effectively guarantee reversal of polarity after each side. If the application of this function generator does not demand that the clipping level E_c be more nearly equal in amplitude to E_o than about $-0.8\,E_o$, then the present circuit will operate reliably; in other words, this generator will easily produce side times five times as long as T_o (the center side time), but the probability that the integrator input polarity will not reverse after every side increases as one attempts to set the clipping level closer to $-E_o$ and obtain longer side times in the output. When the input polarity does not reverse after a side, the output will pass out of the constant peak-to-peak region and take about 0.1 second before the circuit resumes proper operation.

CIRCUIT DIAGRAM

Fig. 10 shows the circuit of the laboratory model of the constant-amplitude random function generator. Fig. 11 shows the cosine shaper and power supply. A general discussion of the circuit operation is given in more detail in Appendix I, while a transistorized generator is discussed in Appendix II.

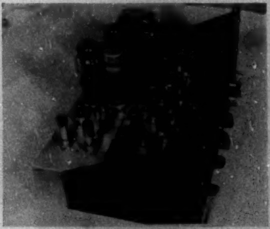


Fig. 14. Laboratory generator

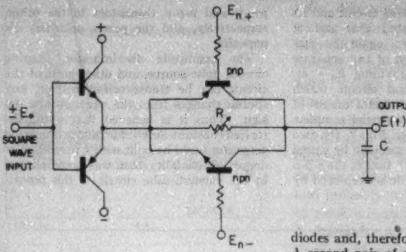


Fig. 15. Transistorized integrator

OSCILLOGRAPHS OF OUTPUT WAVE

Several photographs of an oscilloscope displaying the output wave shape versus time are included, each one a time exposure of many sweeps of the oscilloscope with each sweep beginning at a peak of the wave. Fig. 12 shows the three output modes running at steady state; that is, without any noise voltage. Fig. 13(A) shows three views of the triangle wave using different sweep rates on the oscilloscope; Fig. 13(B) shows three similar views of the cosine wave. In these illustrations the side time distributions can be observed with respect to statistical variations in the choices of side times, the wide range of chosen side times, and the apparently continuous representation of side times when a large number of oscilloscope traces are recorded. Fig. 13(C) and 13(D) differ from the foregoing in that fewer traces are pictured in order to show statistical variations; all of these samples are at the same oscilloscope sweep speed.

Fig. 14 shows the laboratory model with self-contained, regulated power supply.

Appendix I. Detailed Circuit Operation of the Function Generator

The noise source e_s is a *6D4* gas triode (*V*1) operating as a diode in a magnetic field; see circuit of Figs. 10 and 11. In order that the instantaneous noise voltage remain relatively constant during the 10-μsec (microsecond) sampling gate, the noise is bandwidth-limited by a single *RC* low-pass filter with a time constant of about 40 μsec. That is,

$$BW = \frac{10^6}{(2\pi)\,(40)} = 4{,}000 \text{ cps}$$

The filtered noise is then fed through a gain control (*P*-9) and amplified (*V*2) before being put into the dual shaping circuit.

The dual shaping circuit consists of four back-biased diodes. Two diodes (*1N39A*) effect the hard clipping of the noise, one diode in each channel, while a potentiometer (*P*-2) controls the bias on these diodes and, therefore, the clipping level E_c. A second pair of diodes (*1N34A*), each in series with a variable resistance, attenuates the instantaneous signals of the same polarity, in each channel, as those being clipped. A dual potentiometer (*P*-1) controls the amount of attenuation, thus controlling the shaper gain ratio.

The dual sampling circuit is made as follows: Two parallel triodes (*V*3) are operating with a common plate resistor in a normally cut-off state. The shaped noise voltage is direct-coupled to the grids of the triodes, but the grid bias is controlled by two cathode followers (*V*4). When a sample is taken, a 10-μsec pulse is applied to the grid of one of the cathode followers, thus driving the corresponding grid of a sampling triode upward by a fixed amount plus an amount determined by the noise signal. With this arrangement, then, the two shaped noise signals are alternately sampled and fed out at the same point. When the grid biases are properly adjusted (*P*-3, *P*-4) on the sampling triodes, the plate (*TP*-3) will have 100-volt 10-μsec negative pulses appearing.

The voltage on the negative peaks of the sample is fed to the holding capacitor (200 *pf*) through a diode (*V*8), while another diode and pulse generator (*V*7, *V*8) sweep the holding capacitor voltage up to meet the level of the new sample, if the previously held sample was below the new sample. The trigger pulses for both the sampling tubes and the holding capacitor sweeping circuit are obtained by differentiating the two square wave outputs of the amplitude discriminator.

The amplitude discriminator circuit (*V*6) is a bistable multivibrator with a large amount of hysteresis, such that if the input is direct-coupled the multivibrator will flip into one state when the input rises to some specified value, and flop into the other state when the input decays to a second specified value. In this manner, the output of the integrator is contained within constant peak-to-peak limits, since it is the square wave output E_o of the amplitude discriminator circuit which is being integrated. (Of the two out-of-phase square waves available, only one is of the proper phase to be fed to the integrator. Similarly, proper arrangement of the pulses triggering the dual sampler is necessary.)

A potentiometer (*P*-5) controls the effective amplitude of the sampled and held noise compared to the constant amplitude square wave E_o as the two signals are added and power amplified by a cathode follower (*V*10*B*). The center side time T_o is controlled by varying the resistor (*P*-10) at the integrator input. If a large change in T_o is desired, the feedback capacitor (500 *pf*) should be changed. The minimum side time this circuit will generate is about 50×10^{-6} seconds, being limited by responses of the amplifiers.

The integrator operational amplifier (*V*5) contains an internal-positive-feedback loop adjusted by a potentiometer (*P*-12) until a minimum voltage appears at the input grid of the integrator. Two front panel switches are provided. The first (*S*-1) is a function selector, giving either the triangle, cosine, or square wave mode at the output terminals. The second switch (*S*-2) is an output attenuator, the upper position giving a maximum output of about 100 volts, peak-to-peak, while the lower switch position yields about 4 volts. A potentiometer (*P*-11) continuously controls the output level.

The cosine shaping circuit consists of biased diodes arranged to attenuate gradually the peaks of a fixed amplitude triangle wave. Adjustment of the circuit is made by a potentiometer (*P*-8) controlling the input triangle amplitude and a potentiometer (*P*-7) adjusting the level of hard clipping at the wave shape peaks.

Appendix II. Suggestions for a Transistorized Generator

Since the construction of the random function generator, the author has observed some advantages that transistors offer over vacuum tubes in implementing the circuit ideas of this generator. Some qualitative suggestions for transistorized circuitry will briefly be shown in this Appendix, along with the operational advantages which can be achieved. In general, the use of complementary symmetry with transistors enables the design of circuits balanced about ground potential in a push-pull, class *B* fashion.

An integrator is easily made by connecting a capacitor to the collectors of two transistors in parallel, one a p-n-p and the other an n-p-n, as shown in Fig. 15. In this configuration, one of the high-output impedance transistors will feed a current into the capacitor for integration of signals of one polarity, while the other transistor will be used to draw a current from the capacitor for integration of signals of the opposite polarity. The most attractive feature of this integrator for use in the function generator is obtained from the nature of the transistors and the integrator input arrangement, which is such that the input polarity can be reversed after each side with certainty, no matter what amplitude of noise signal is being integrated, as mentioned previously.

The input to the emitters is the square wave output of an amplitude discriminator sensing the integrator output $E(t)$; see Fig. 5. When the square wave is positive, the noise sample E_n+ will determine the constant current flowing in the capacitor, while E_n- should be equal to $+E_o$ during this time in order for the current through the n-p-n transistor to be negligible and to have the same polarity as the current through the p-n-p transistor. Should the amplifier providing the noise sample E_n+ present a sample more positive than E_o, the transistor becomes a perfect clipper in that such a noise sample cannot possibly deliver a current of the wrong polarity to the integrating capacitor. Conversely, during integration of $-E_0$ and its corresponding noise sample E_n-, E_n+ should always be equal to $-E_0$. The resistor R determines the longest side time that can be produced. Note that the output peak amplitude will necessarily be less than E_0.

In view of the previous discussion, the noise sampling circuit need only to provide a shaped noise signal with no clipping, thus eliminating any exacting requirements on the accuracy of the sampling circuit and its amplifiers. It is suggested that use of complementary symmetry again be employed in the construction of two separate sampling circuits, one providing E_n- alternately with the second circuit which supplies E_n+. Since the useful ranges of output voltage of these two noise samplers differ only in polarity, see Fig. 7, the two sampling circuits would especially be suited by complementary design; that is, the n-p-n and p-n-p transistors in one are replaced by p-n-p and n-p-n transistors in the other, respectively, and the power polarities are opposite.

The amplitude discriminator, shaping circuit, noise source, and other parts of the circuit can be transistorized without any special changes from the vacuum-tube design. Thus it is believed that a transistorized constant-amplitude random-function generator could be built with a much higher degree of reliability than was accomplished in the vacuum-tube circuit of this report.

Current-Balancing Reactors for Semiconductor Rectifiers

I. K. DORTORT
MEMBER AIEE

THE ADVANTAGES of the semiconductor diodes in certain applications of d-c power have led to the design and construction of rectifier units of very high current ratings, employing many diodes in parallel on each phase of the rectifier transformer. As with mercury-arc rectifiers with more than one anode connected to each phase, the designer is faced with the problem of current unbalance between parallel elements.

Although the major causes for current unbalance in the two types of rectifiers are quite different from each other, balancing reactors will serve in either case to force proper current distribution. While these reactors resemble each other physically, their modes of operation are different. Moreover, the large number of semiconductor diodes often connected in parallel will generally dictate a different arrangement. It is felt that a brief description of the operation of balancing reactor with mercury-arc rectifiers will be useful as a preliminary to the main subject of this paper.

Anode-Balancing Reactors

In any well designed and properly maintained mercury-arc rectifier, the arc-drop characteristics of the individual anodes are so nearly alike that artificial means of absorbing the voltage difference between them is seldom required. The resistance and inductance of symmetrically arranged anode leads, the flux produced in the iron anode shaft and in the tank wall or anode shield are sufficient to maintain good current balance once it has been established in any conduction period. However, the ignition voltage required to start the arc is often considerably higher than the arc drop needed to sustain the peak current during the rest of the conduction period. Since the two anodes are tied together, the first one to fire will pull down the voltage of the parallel, connected anode to a point below its ignition voltage, and may prevent it from picking up at all during the rest of the cycle. Two parallel anodes will not pick up simultaneously except by rare coincidence. Other disturbing factors may exist, such as the erratic behavior of ignitors in ignitron rectifiers.

In any case, conduction of an anode does not take place immediately after the commutating voltage becomes positive, but some finite time later as at a in Fig. 1, when the ignition voltage E_{pf1} is reached. The anode-to-cathode voltage then drops sharply to the normal arc drop, the difference between the anode and commutating voltage being taken up by the commutating reactance, as indicated by the total, crossed area.

If two anodes are paralleled through a center-tapped anode reactor, a proportional part of the reactance drop occurs across that half of the reactor connected to the anode which picked up first, raising the voltage of the center tap above the anode by that amount. The same voltage appears across the other half of the reactor, raising the voltage of the idle anode above the conducting anode by the full voltage of the reactor. When the full reactor voltage attains the difference between the ignition voltage of the idle anode and the arc drop of the conducting anode, the second anode will fire.

Fig. 1(B) illustrates the current-balancing action of the anode reactor. The curved line i_c represents the commutation current build up of the incoming phase, if conduction starts at the voltage transition point. It is not affected by the number of anodes firing, as long as the commutating impedance is not materially changed. When firing is delayed, for whatever reason, the current build up follows the same curve, shifted downward, parallel to itself, so that it starts at zero. This is illustrated by the heavy portion of curve i_c' in Fig. 1(B).

When the second anode is forced to fire at time b, there is some change of impedance, since the balancing reactor is effectively short-circuited. When the change is small, both anode currents will follow curves parallel to $i_c/2$, displaced so that their sum will always be equal to i_c'. Both anode currents continue to rise to the end of the commutating period or angle of overlap u, at which time the preceding phase current has dropped to zero and the new phase carries the total load current of the commutating group.

From time b until the end of u the two anode currents rise parallel to each other, and then maintain their difference substantially constant, to the end of the conduction period. The maximum difference to be allowed under adverse conditions determines the latest time b when the second anode must pick up.

The reactor must be designed to produce the excess voltage $E_{pf2}-E'_{a1}$ on anode 2 no later than point b in the cycle. The cross-hatched area represents the minimum volt-seconds that must be built into the reactor. However, no additional volt-seconds are required beyond b, except as design margin. For the rest of the

---

Paper 58-219, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 8, 1957; made available for printing November 25, 1957.

I. K. DORTORT is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

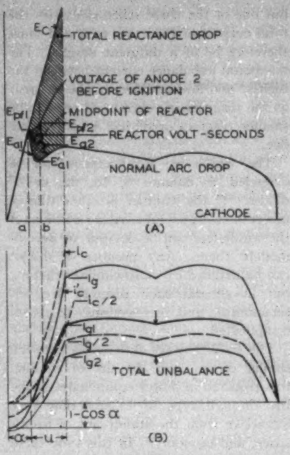


Fig. 1. A—Voltage relationships on mercury-arc anode balancing reactor; E_{pf1}, ignition voltage of anode 1 at time a; E_{a1}, arc drop of anode 1; E_{a1}', arc-drop of anode 1 at time b when anode 2 picks up; E_{pf2}, ignition voltage of anode 2; E_{a2}, arc drop of anode 2. B—Current division in anode reactor; I_c, total commutating current with no delay $=(\sqrt{2}\ E_s/2X_c)(1-\cos \omega t)$; I_c', total commutating current with first anode delayed angle α, $=(\sqrt{2}\ E_s/2X_c)(\cos \alpha-\cos \omega t)$; I_g, total load current on both anodes; I_{g1}, load current on anode 1; I_{g2}, load current on anode 2; α, angle of delay of anode 1 for all causes; u, angle of overlap

conduction period the two anodes are at almost the same potential, and the balancing reactor serves no further purpose.

This principle can be extended to the design of reactors for paralleling three, or even four, anodes. Such rectifiers have been built in the past, but modern practice limits the number of parallel anodes to two. The balancing reactor is therefore a relatively simple device.

Semiconductor Rectifiers

In contrast to the mercury arc, a semiconductor diode has nothing which corresponds to the ignition voltage. Appreciable conduction begins at a relatively low threshold voltage, but, as can be seen in Fig. 2, the current-voltage curve quickly becomes very steep. These curves were drawn for a particular type and class of diodes, and much steeper characteristics have been observed. Because of this steepness, even a few millivolts' difference between two diodes operating at the same current may mean a large difference of current when they operate in parallel, and therefore at the same voltage. This is clearly demonstrated by the equal-current and equal-voltage lines in Fig. 2.

Although some surprisingly heavy current diodes have been produced, the current ratings of large rectifiers require, and probably will require for some time to come, a considerable number of diodes in parallel. Although the number depends on the ratings of cells and the rating of the complete rectifier, as many as 20 or more diodes in parallel may be required. Larger diodes undoubtedly will be developed and produced economically some day, but it is expected that the problem of paralleling will exist for a long time.

In the application of heavy-current selenium rectifiers for plating and similar applications, it has not been uncommon practice to provide a ballast resistor in series with each stack, primarily to prevent a new replacement stack from hogging the load and destroying itself. The use of germanium and silicon rectifiers in the field of power makes necessary a re-examination of the paralleling problem and a re-evaluation of the solutions available. The treatment of these has been varied and ranges from completely natural current distribution to forced current balancing of high accuracy. The following list, with annotations, offers a quick review of the methods used, their merits and limitations.

1. *Natural current distribution.* Only effective to any reasonable extent when several diodes are connected in series, and the series strings are paralleled. Statistical variation offers some balancing effect, depending on the number of diodes in series. Unbalance tends to worsen with increasing load and temperature.

2. *Forward-drop matching.* Use of matched diodes requires careful coding, and replacement of diodes with new diodes in the same code group. Characteristics matched at one load may deviate with changes of load and temperature and possibly with time. If the diodes are not also matched on inverse characteristics, current distribution will also vary, to some extent, with variations of applied voltage.

3. *Ballast resistors.* Effective only when the resistance drop is made a significant part of the total drop. Losses are increased in proportion. When used with coded diodes the resistance drop of protective fuses may assure sufficient balancing.

4. *Derating factors.* In any of the preceding three methods, derating factors are generally applied as some function of the number of diodes in parallel, for purposes of reducing the percentage of failures caused by overloaded cells. Discussed in more detail subsequently.

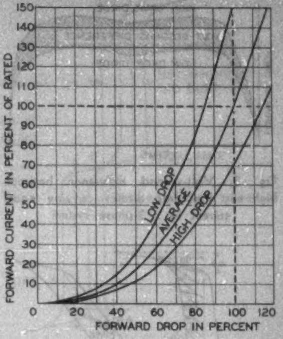


Fig. 2. Peak forward current plotted against peak forward voltage drop for a particular type of diode, based on the average

5. *Current-balancing reactors.* Discussed in detail subsequently.

Resistors, in varying degree, reduce the potentially high efficiency of semiconductor rectifiers. In low-voltage applications, where the total power is small despite high currents, this may not be very serious. Improved current balance by means of matched diodes is burdensome, and will remain so until improved manufacturing techniques provide better control of the forward drop than at present. The use of drastic derating factors is an expensive necessity in the absence of other means of current balancing. For holding down the percentage of cell failures, factors have been proposed which reduce the rating of each additional cell, in arithmetical or geometrical progression: In some cases the point of diminishing returns is approached very rapidly, and a further addition of cells does not increase sufficiently the total rating of the rectifier to justify them.

In view of these circumstances the use of balancing reactors deserves careful consideration, not only from the standpoint to technical correctness, but as a matter of economic advantage. In this respect we must not overlook the importance of continuity of service and cost of maintenance.

THEORY OF OPERATION

The current-voltage characteristics of similar commercial diodes fall on either side of the average characteristic, and are roughly parallel to each other. Therefore, the difference in voltage drop between two diodes, paralleled through balancing means so as to carry equal currents, will be substantially constant

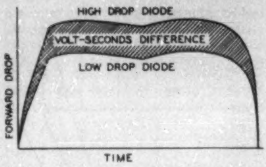


Fig. 3. Volt-second difference between high and low drop diodes carrying equal current in a 3-phase system

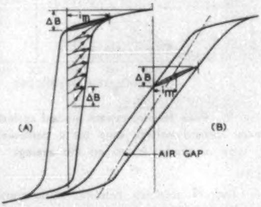


Fig. 4. A—Magnetic behavior of current-balancing reactor without air gap or external flux reset means. B—Same reactor with small air gap. ΔB, flux change required to absorb volt-seconds of Fig. 3. i_m, magnetizing current averaged over conducting period

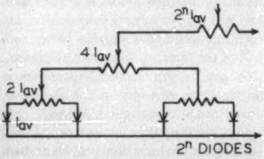


Fig. 5. Whiffletree arrangement of balancing reactors

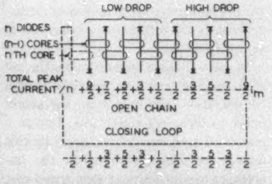

LOW DROP
HIGH DROP
n DIODES
(n-1) CORES
n TH CORE
TOTAL PEAK CURRENT/n $+\frac{9}{2}+\frac{7}{2}+\frac{5}{2}+\frac{3}{2}+\frac{1}{2}-\frac{1}{2}-\frac{3}{2}-\frac{5}{2}-\frac{7}{2}-\frac{9}{2}i_m$
OPEN CHAIN
CLOSING LOOP
$-\frac{1}{2}+\frac{1}{2}+\frac{3}{2}+\frac{5}{2}+\frac{3}{2}+\frac{1}{2}-\frac{1}{2}-\frac{3}{2}-\frac{5}{2}-\frac{3}{2}-\frac{1}{2}$

Fig. 6. Chain of reactors showing current unbalance with worst distribution of diodes; maximum difference of peak currents with open chain is $(n-1)i_m$. Maximum difference of peak currents with closed chain is $(n/2)i_m$

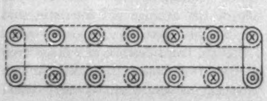


Fig. 7. Method of closing chain of reactors

and in the same sense during most of the conduction period. Fig. 3 illustrates this mode of operation, and the shaded area represents the volt-seconds that must be absorbed by the balancing reactor. This area translated into flux linkages fixes the minimum cross section of iron and the number of turns of the reactor. The unbalance between the two diodes at any instant is the instantaneous magnetizing current required to produce the flux change represented by the volt-second area up to that time.

The degree of balance therefore depends on the quality of iron and of structure used in the reactor. It is to be expected that, in the simplest form of reactor, without any means of flux reversal or resetting, the core will be excited beyond, and return to, its residual flux density at the end of each conduction period. Even if the voltage difference between the two diodes is very small, the volt-seconds created in successive cycles are cumulative, and starting at zero, will soon reach the residual flux density. This is illustrated in Fig. 4(A) which is based on a typical, oriented transformer iron core without air gap. Fig. 4(B) illustrates the behavior of a similar core built with a small air gap. The magnetizing current is increased over the full range of flux density, but the residual flux is so much lower that the average current over the useful range is less than in the reactor without any air gap. Unfortunately, the requirements of low magnetizing current and low residual flux work in opposite directions.

For a reasonably close approximation an average can be taken of the magnetizing current over the conduction period, for the range of flux densities over which the reactor will be working, and this average can be applied to the peak diode currents for a determination of the amount of unbalance. Since the magnetizing current can affect the diode current only during the conduction period, the unbalance expressed in average current over the full cycle is only one-third of the magnetizing current when operating in a 3-phase system. The unbalance of the diode rms currents is the magnetizing current divided by $\sqrt{3}$.

Reactor Systems

The simplest system of balancing reactors is obtained when only two diodes are to be paralleled, in which case a single, center-tapped reactor fulfills the requirements. When four diodes are to be paralleled, a whiffletree arrangement, as illustrated in Fig. 5, can be employed. This is still a fairly simple arrangement, but one of the three reactors carries the total current of the other two, and must therefore be of a different design. The maximum unbalance current between the highest and lowest diode can be as great as the magnetizing current of one of the smaller reactors or of half of the larger one.

The whiffletree arrangement can be extended to balance 8, 16, 32, or 2^n diodes. If the number is other than a power of two, unsymmetrical variations of the whiffletree can be devised to accommodate them. Any number of diodes can be handled by multiwinding reactors, but in general such reactors are uneconomical and connections to them complicated.

The small voltage differences between similar diodes, and the size of conductors required in power equipments, make the through type of construction very attractive from the standpoint of losses, space, and economy. In this type there are only two conductors in the core window; they carry the currents of two diodes in opposite directions. The difference between these currents is the net magnetizing current acting on the core.

With such reactors, it is relatively simple to balance each diode against its immediate neighbor in "chain" fashion, as shown in Fig. 6. The greatest unbalance will occur if all the diodes in one-half of the chain are in the highest voltage-drop class, and all those in the other half in the lowest voltage-drop class. Since adjacent conductors thread the same core in opposite directions, the greatest instantaneous unbalance between them cannot exceed the instantaneous magnetizing current of that core, i_m. Each diode will, therefore, carry a current equal to the average current of all the diodes, plus or minus some factor of magnetizing current, as indicated in Fig. 6 for 10 diodes. In the case of "n"

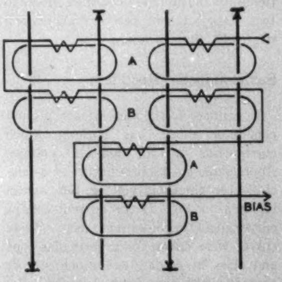


Fig. 8. Two-way biased reactors

VoL.
Pt. 1
195

diodes in an open chain, the maximum deviation from the average, taken over a full cycle, is $\Delta I_{av}=(1/6)(n-1)i_m$ in a 3-phase system.

If the chain is closed on itself by adding one more core ahead of the first diode and threading through that core the lead of the last diode, balancing voltages are transmitted in two directions at the same time. This produces the current distribution shown in the second line of Fig. 6. The maximum unbalance is reduced by approximately one-half, and $\Delta I_{av}=(1/12)ni_m$.

This method can be used when not too many diodes are connected in parallel. When large numbers of diodes are paralleled, the closing lead, shown as a dotted line in Fig. 6, becomes undesirably long. A better way which eliminates this long lead and offers other structural advantages, is to double the chain back on itself, as shown in Fig. 7. All diode leads are short and conveniently arranged.

Design of Reactors

As previously mentioned, oriented, magnetic-steel cores, either rolled or punched, without air gaps, would have the smallest magnetizing current and highest maximum flux densities, and also, unfortunately, the highest residual flux densities. When using only air gaps or lower quality steels to reduce the residual flux, one must make a careful comprise to obtain the best balance with the least material and space.

There are many ways of reducing the residual flux electrically, but the greater complexity of the system needed for this may not be justified. For instance, each reactor may be made up of two cores, each core oppositely biased, as in Fig. 8. The magnetic diagram for this system is given in Fig. 9; it is obvious that the system comes to rest at point *O* at the end of each conducting period. The bias ampere-turns must be slightly greater than the magnetizing current of each core component, and the net magnetizing or unbalance current will therefore be slightly greater than twice the magnetizing current of the core without bias. Nevertheless, since the core can be made of the best magnetic material, without an air gap, the unbalance current will still be very small. It should be noted, however, that one core will be idle most of the time, yet must be furnished unless the direction of unbalance can be predicted and maintained. It becomes a matter of economy to decide whether two small cores with bias windings are more desirable than a single core containing more iron but having no windings or bias supply. The same reasoning must be applied to other electrical means of reducing the residual flux of single cores, but these again involve secondary windings and a certain amount of auxiliary apparatus.

Up to the present time it appears that even though reactors without flux-resetting means make poor use of the iron, the relative cheapness of the material and labor involved make the simple through-type reactor the most economical. Preformed, wound cores with one or more air gaps are commercially available and offer a good solution to the problem. Another very satisfactory and economical solution is a stack of punched laminations, as illustrated in Fig. 10(A). Cold-rolled transformer steel, oriented lengthwise, without air gap, has proved surprisingly satisfactory. The cross orientation around the ends of the slots reduces the residual flux to a tolerable level. It has been found expedient, however, to slit some of the laminations lengthwise, as shown in Fig. 10(B), so as to provide a partial air gap.

Test Results

The strip-type structure of Fig. 10(A), without closure, was tested with 12 150-ampere 94-PIV (peak inverse volts) germanium cells, loaded in a half-wave circuit to 1,570 amperes d-c total. The cells were not matched, and were placed at random. The results of this test are shown in Fig. 11. Curve *A* shows the current distribution among diodes without any balancing iron; the maximum deviation amounts to 65% from the average diode current. Curve *B* shows the current distribution with balancing iron having no air gap; the maximum deviation is 7.5% from average. Curve *C* gives the distribution with a percentage of the laminations slit longitudinally, as in Fig. 10(B), to provide a partial air gap; the maximum deviation was reduced to 4.0%.

Conclusion

Especially when it is desired to connect both ends of the diodes to paralleling busses so as to take advantage of liquid-cooled busses, and of the voltage divider system described in an earlier paper, it becomes important to develop a simple balancing reactor which will not greatly increase the space requirements, the complexity, and the cost of a large rectifier. Fig. 12 shows the compact and convenient layout of one phase of a 5,000-ampere section of a 10,000-ampere 300-volt silicon rectifier using two series-connected banks of 20 parallel

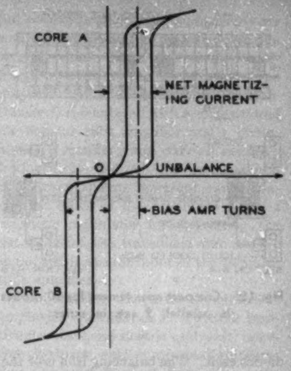


Fig. 9. Magnetic diagram of 2-way biased reactors

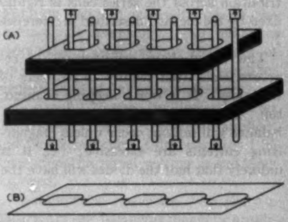


Fig. 10. A—Punched-strip open-chain reactor. B—Method of producing partial air gap

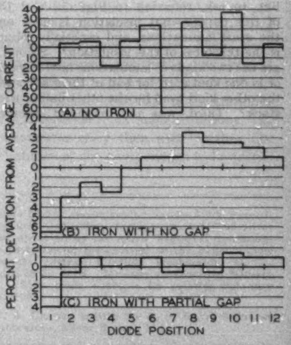


Fig. 11. Current distribution among 12 parallel diodes with: A—no iron, B—iron with no air gap, C—iron with partial air gap

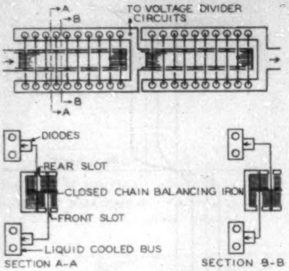


Fig. 12. Compact arrangement for 20 diodes in parallel, 2 sets in series

diodes each. The balancing iron is of the form shown in Fig. 13. The busses are liquid-cooled, and the diodes mounted on them can be replaced without breaking into the cooling system. Voltage distribution across the series cells is rigidly fixed by the method described in reference 1.

The closed-chain arrangement of reactors lends itself readily to rectifiers of this class. It is worth keeping in mind, too, that the maximum predicted unbalance will seldom occur unless magnetizing currents are excessive; i.e., it is unlikely that half the diodes will have the highest voltage drop, and half the lowest voltage drop, of a class, and that they will be nicely segregated at opposite ends of the chain.

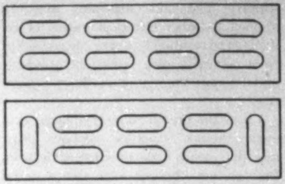


Fig. 13. Punched-strip closed-chain reactor

There has been much work in various places and along various lines to overcome the disadvantages of dissimilar diode characteristics. The application of drastic derating factors to compensate for the poor distribution of voltage and current even among coded diodes must be considered based on a wasteful philosophy of necessity, and is not justified. In small rectifier units using few cells, current and voltage balancing are both easy and inexpensive to accomplish. In large units the need and advantages are obvious:

1. Operation of each diode under the conditions for which it was intended is assured.
2. The cost is offset by the saving in cells which are otherwise required by derating.
3. The added space is offset by the space requirement of extra cells.
4. The operation of diodes at rated current and voltage should assure longer life, and result in savings for the user, the equipment manufacturer, and the diode manufacturer.
5. Operation of diodes within their ratings should improve reliability and availability of the equipment, and reduce maintenance problems.
6. The complete elimination of coding, or at least the broadening of classifications, is important to the user in the reduction of stock items, the chances or error, and in availability. It is hardly less important to the manufacturer.

As the technology of semiconductor diodes improves, and larger ratings with closer tolerances become available, the need for balancing may be reduced. It seems likely, however, that unit ratings will also go up, still requiring banks of cells in parallel, and possibly in series. The greater cost and percentage of unit rating invested in each large cell will probably justify balancing arrangements, even though manufacturing tolerances are greatly improved.

Reference

1. A NEW VOLTAGE-DIVIDER CIRCUIT FOR SEMICONDUCTOR RECTIFIERS, I. K. Dortort. *AIEE Transactions*, vol. 76, pt. I, July 1957, pp. 356–58.

◆

Discussion

A. A. Halacsy Federal Pacific Electric Company, Newark, N. J.): I would like to ask, referring to Figs. 12 and 13 of this paper, what insulation is used on the steel laminations, and especially what is the thickness of this insulation? Second, are the two kinds of laminations interleaved, or is one kind used for half of the stack and the other kind used for the other half of the stack? Third, was any magnetic leakage observed from one kind of lamination to the other kind of lamination, which would result in a closed magnetic path around one conductor only, or was this leakage so small that it could not be observed at all?

I. K. Dortort: Mill-coated transformer steel was used in the strip reactors of Figs. 10, 12, and 13. No insulation was added between laminations.

The two kinds of laminations are purposely arranged in two separate stacks, one above the other as shown in Fig. 10, and separated by a strip of insulating material. Any appreciable interchange of flux between the two stacks must take place across the insulating spacer as well as the gaps between laminations close to the separator.

The maximum magnetomotive force in the iron around the complete loop is very small even at maximum flux density. Even if all the available ampere-turns were to appear across the effective air gap, the flux density in the air gap would be negligibly small. For this reason no attempt was made to observe or measure this "leakage flux."

Magnetic-Amplifier-Operated Ink Recorders

W. A. GEYGER
MEMBER AIEE

Synopsis: The purpose of this paper is twofold. It is intended to describe in a summarizing way the specific properties of magnetic amplifiers with regard to instrumentation. It is also particularly concerned with the application of self-balancing magnetic-amplifier techniques in the realm of moving-coil ink recorders and ink oscillographs.

The paper, therefore, presents detailed information on design and actual performance of high-precision magnetic instrument amplifiers exhibiting unusual stability of operation, linearity of the output-input characteristic, and freedom from zero drift. Such d-c amplifiers offer the general advantages of elimination of movable parts and vacuum-tube components, sturdiness and reliability, and almost unlimited life. They are, therefore, very practical for long-time reliability testing of various circuit components such as high-sensitivity relays, where absolute reliability of the d-c amplifier, which operates the ink recorder, is of prime importance. Actually, amplifier-input signal voltages of the order of 10 to 100 mv (millivolts) and amplifier-input signal currents of the order of 2 to 50 μa (microamperes) have been used successfully for operating standard-type moving-coil ink recorders and ink oscillographs without impairing the accuracy of measurements.

Magnetic-amplifier-operated ink recorders are also well suited to perform mixing of signals; i.e., algebraic summation of signal voltages and signal currents. Multiplication of signals may be performed by means of two thermocouples, one of which is heated by the algebraic sum of two derived currents, and the other by their difference. Another possibility to perform multiplication consists in replacing the two thermocouples by two magnetic-amplifier multiplying circuits, which act as square-low means and provide the square term of the algebraic sum of two derived currents, and the square term of their difference, respectively.

A Brush-type ink oscillograph in conjunction with a 400-cycle-operated 2-stage or 3-stage magnetic amplifier has been used for recording 30-cycle signals with attenuation not exceeding 5%. The paper presents samples of typical ink oscillograms which have been obtained by recording variable-frequency sinusoidal-waveshape and square-wave signals. Through the use of a very large amount of negative feedback in self-balancing magnetic amplifiers, extreme stability with regard to zero drift and constant gain is assured in various ink-recorder applications.

THE HISTORICAL development of the magnetic-amplifier art[1,2] makes it evident that there is a very close linkage between saturable-reactor devices and electrical measurements. Pertaining patents, scientific papers, and technical reports may be divided into main groups. The first group deals with various methods for performing measurements on magnetic amplifiers and similar devices, to obtain information concerning basic principles of operation, steady-state characteristics, and dynamic performance of such amplifiers. The second group is concerned with the application of magnetic amplifiers in instrumentation; e.g., measurement of large direct currents and high direct voltages, extending the sensitivity of electric instruments, and operation of high-speed self-balancing potentiometers and bridge networks with magnetic servo-amplifiers.

In order to replace the very sensitive and delicate instruments, as used in connection with various forms of transducers (thermocouples, photoelectric cells, resistance-thermometer circuits, resistance strain gages, etc.) by rugged indicating and recording instruments, a d-c amplifier with adequate zero stability and stable, linear characteristic must be interposed between the transducer and the instruments. Such d-c amplifiers are particularly suitable for extending the sensitivity of electrical multirange instruments and moving-coil ink recorders. They may also be used in connection with phase-sensitive rectifiers for the measurement of various a-c quantities, for example, in a-c potentiometer circuits and a-c bridge networks. Voltage dividers and shunts may be provided to extend the voltage and current ranges of such arrangements.

In instrument applications, the magnetic amplifier is used either as an "all-magnetic" device or as a magnetic modulator in combination with vacuum tubes or transistors. In the first case, single-stage or multistage magnetic-amplifier push-pull circuits utilizing over-all negative feedback for stabilization have been successfully used in various forms. Although speed of response is generally of minor importance in these applications, the basic requirements of adequate zero stability and stable, linear output-input characteristic (accuracy about 1 to 2% of full-scale indication of the instrument) are sometimes difficult to satisfy. In the second case, either a fundamental-frequency-type or a second-harmonic-type magnetic amplifier (or flux-gate magnetometer[3]) is combined with a conventional a-c vacuum-tube amplifier and a phase-sensitive rectifier; and over-all negative feedback is utilized for stabilization. Such arrangements represent carrier-frequency types of d-c amplifiers using a magnetic modulator for converting the d-c signal of the transducer into an a-c signal applied to the input of the vacuum-tube amplifier.

Self-balancing magnetic amplifiers[4–6] employing combination of magnetic positive feedback and electric (galvanic) negative feedback exhibit unusual stability, linearity, and freedom from zero drift, which are requirements of prime importance in electric-instrument applications. These amplifiers offer the possibility of reducing actual response time to its minimum value of one-half cycle of the power-supply frequency. For instance, when using a power-supply frequency of 400 cps (cycles per second), the amplifier reaches 100% of its final steady-state output current 1.25 msec (milliseconds) after the transient signal voltage is applied; and the output current drops to zero 1.25 msec after the signal voltage is suddenly removed.

This paper is concerned with the application of self-balancing magnetic amplifier technique in the realm of moving-coil ink recorders and ink oscillographs. Such high-precision "magnetic instrument amplifiers" are very practical for long-time reliability testing of various circuit components; e.g., high-sensitivity relays, where absolute reliability of the d-c amplifier, which operates the ink recorder, is of prime importance. There have been even some cases, where the operation of the components to be tested has proved to be more reliable than that of the amplifier of the recorder. In such situations, the investigator may then ask whether the reliability of the components to be tested

---

Paper 58-541, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28–30, 1958. Manuscript submitted January 20, 1958; made available for printing February 21, 1958.

W. A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

The author wishes to express his appreciation to B. T. Hooper, A. D. Krall, and W. Raskin for their helpful co-operation in conducting the oscillographic investigations. Indebtedness is also acknowledged to W. P. Christoph for valuable discussions regarding multiplying circuits and various applications of magnetic-amplifier-operated ink recorders.

or that of the amplifier, which drives the moving-coil ink recorder or ink oscillograph is being tested.

Recently, a considerable part of the author's activity has been concentrated on special forms of magnetic amplifiers for standard-type moving-coil ink recorders and ink oscillographs. This work has been directed toward the operation of Esterline-Angus and Brush recorders with self-balancing magnetic amplifiers working under either voltage-balance or current-balance conditions. With regard to various special instrumentation problems, these d-c amplifiers have proved to be powerful tools for providing extremely stable magnetic-amplifier-operated ink recorders. Such recorders are also well suited to perform algebraic summation and multiplication of signal voltages or signal currents by means of thermocouples or magnetic-amplifier multiplying circuits acting as square-low means, respectively.

An important part of these investigations has been concerned with the combination of a Brush recorder with a 400-cycle-operated multistage magnetic instrument amplifier, each stage of which has a response time of one-half cycle of the power-supply frequency (1.25 msec). Variable-frequency sinusoidal-waveshape and square-wave signals have been employed for testing the dynamic performance of these arrangements. Typical ink oscillograms proved that it is possible to perform in this way recording of 30-cycle signals with attenuation not exceeding 5%.

When employing a conventional 1-ma (milliampere) 1,500-ohm Esterline-Angus recorder, amplifier-input signal voltages of the order of 10 to 100 mv and amplifier-input signal currents of the order of 2 to 50 μa have been used for operation of the recorder without impairing accuracy of measurements. In this case, the response time of the 60-cycle-operated magnetic instrument amplifier is of the order of 100 msec, or even more. Through the use of a very large amount of negative feedback in this amplifier, extreme stability with regard to zero drift and constant gain is assured in various ink-recorder applications.

Basic Considerations on Magnetic Instrument Amplifiers

General

The relationship between power gain, response time, and stability of a magnetic amplifier is such that if one property is accentuated, it is always at the expense of the the other properties. Recognition of this fact is particularly important with regard to magnetic amplifiers to be used in instrumentation.

For example, in conventional external-feedback or self-saturating magnetic-amplifier circuits, the dynamic power gain D_P, i.e., the ratio of usable power gain K_P over control time constant T_C (sometimes called the "figure of merit"[7]), is given by the equation

$$D_P = \frac{K_P}{T_C} = 4f_P \frac{R_L}{R_O} K_{AT} k_f^2 \tag{1}$$

or

$$D_P = 4\frac{R_L}{R_O} K_{AT} k_f^2 \text{ per cycle} \tag{2}$$

where f_P is the power-supply frequency, R_L is the load resistance, R_O is the total resistance of the load circuit, K_{AT} is the ampere-turn gain, and k_f is the form factor; i.e., the ratio between rms value and average value of the current in the load circuit. On the other hand, in these conventional circuits, the drift is proportionally linear to the actual value of ampere-turn gain K_{AT}. Obviously, the drift can be reduced by reducing K_{AT}. However, because the ratio of gain reduction over drift reduction is constant, the gain will be reduced at the same rate.[8]

In self-balancing magnetic amplifiers, a very significant reduction of drift is obtained by reducing the effective (total) feedback factor,[9] which is a measure for the stability of operation. This has been achieved by combination of the dynamic gain-increasing effect of magnetic positive feedback and the drift-reducing effect of electric negative feedback, resistance-coupled into the control-circuit loop.

The Evolution of the Self-Balancing Magnetic Amplifier

The ordinary saturable reactor is characterized by the fact that, over certain ranges, the average value of load current is substantially independent of large variations in magnitude, frequency, and waveshape of the power-supply voltage, and considerable changes in load resistance. Therefore, it is possible to employ ordinary saturable reactors with high-permeability rectangular-hysteresis-loop core material as d-c instrument transformers for measurement of large direct currents and high direct voltages. This function is based upon the current-transformer character ("law of equal ampere-turns"[10]) which is the result of inherent negative-feedback properties ("intrinsic" negative feedback) of the ordinary saturable reactor having either series- or parallel-connected a-c load windings. Moderate power amplifications can be obtained with such simple arrangements when the load resistance R_L is properly matched to the impedance values of the saturable-reactor output circuit so that the output power P_L is greater than the input power P_C, representing the actual losses in the d-c control circuit. The value of the power gain $K_P = P_L/P_C$ (about 10 to 100) depends on the design and size of the saturable-reactor elements preferably having spirally wound tape cores with multilayer toroidal windings.

In order to obtain much higher values of amplification ($K_P = 1{,}000$ to 10,000, or more), it is common practice to apply a very large amount of positive "extrinsic" feedback by means of external-feedback or self-saturating circuits for reducing or nearly nullifying the intrinsic negative feedback of the ordinary saturable reactor. By connecting a high resistance (sometimes termed "forcing resistor") in series with the d-c control windings, it is possible to obtain well-known static and dynamic performance characteristics of such extrinsic-feedback types of magnetic amplifiers, as expressed, e.g., by ampere-turn gain, power gain, control time constant, and dynamic power gain (figure of merit).

Although conventional forms of external-feedback or self-saturating magnetic-amplifier circuits have proved very suitable for many applications, these circuits generally do not meet the severe requirements where very high power gain (high sensitivity), extreme stability with regard to zero drift and constant gain, and high speed of response are of paramount importance. When applying quantity-production techniques, these performance characteristics of the magnetic amplifier must be secured with a minimum of core-matching and component selection.

The aim of the research and development work on magnetic amplifiers of the self-balancing potentiometer type[4] consisted in combining the inherent extremely stable operating conditions of the ordinary saturable reactor, as used in d-c instrument transformers[11,12] with the unique features of high-precision self-balancing potentiometer devices and negative-feedback amplifiers. In these self-balancing magnetic amplifiers, extrinsic magnetic positive feedback (external feedback or self-saturation) is provided for completely nullifying the intrinsic negative feedback of the ordinary saturable reactor (infinite internal gain or "critical regeneration"[4–6]). Additionally, extrinsic electric negative feedback (resistance-coupled into the control-circuit loop) is applied to obtain very stable operating conditions of the amplifier, similar to those of d-c instrument transformers. In this way, by combining the dynamic gain-increasing effect of magnetic positive feedback with the drift-reducing effect of electric negative feedback, a very significant reduction of drift is obtained, because the effective feedback factor of the self-balancing magnetic amplifier is considerably reduced.[6] Operating under "voltage-balance" or "current-balance" conditions of the control circuit, this amplifier demands practically no energy from the d-c signal source; e.g., a thermocouple or a phase-sensitive rectifier circuit. Because the average value of the control-winding current is substantially zero for the entire working range of the amplifier, the output-input characteristic is linear (the linearity error is of the order of 0.1% or less). Over certain ranges, large variations in magnitude, frequency, and waveshape of power-supply voltage, load resistance, and control-circuit resistance have no appreciable effect upon the transconductance, voltage ratio, current ratio, or transresistance of the amplifier. The very large amount of electric negative feedback used makes the gain of the amplifier essentially independent of all factors except the accuracy of the resistors in the negative-feedback network.

Self-balancing magnetic amplifiers operate under extremely stable conditions with a large over-all power gain (about 1,000 to 10,000 in a single-stage design) and a high speed of response (the response time may be reduced to its minimum value of one-half cycle of the power-supply frequency). On the other hand, for special applications, the transient-response characteristics may be altered materially through the use of integrating circuit elements in the degeneration network of the amplifier. Of course, in this case too, the relationship between power gain, response time, and stability of operation is such that if one property is accentuated, it is always at the expense of the other properties. Results of various experimental investigations have shown that the actual behavior of self-balancing magnetic amplifiers approximates very closely corresponding characteristics of an ordinary saturable reactor with stabilizing negative-feedback properties (intrinsic negative feedback).

Operation of Magnetic Amplifiers with Magnetic Voltage Stabilizers

ELEMENTARY IDEAS

The significant fact that the output-input characteristic of an ordinary saturable reactor and that of a properly designed self-balancing magnetic amplifier is substantially independent of large changes in voltage E_P and frequency f_P of the a-c power supply means that actual value of magnetic-flux density in the cores, $B = \text{constant} \times E_P/f_P$, may change within wide limits without appreciably affecting the performance of such arrangements. However, when the problem consists in providing very small input-voltage ranges; e.g., 0–5 mv or 0 to ±2.5 mv, then flux-density variations, as caused by voltage fluctuations of ±10% and frequency variations of ±5%, may produce excessively large zero-drift errors and changes of the output-input characteristic. In such cases, application of additional means for eliminating undesirable changes of actual value of flux density in the magnetic-amplifier cores has proved to be necessary. In 1940, the author showed[13] that the magnetic voltage stabilizer,[14] a cousin of the magnetic amplifier, is extremely suitable for this purpose. Magnetic voltage stabilizers consisting of transformer-like structures and capacitors (ferroresonant circuits) have no moving parts of vacuum tubes and can perform and added function of a step-up or step-down transformer, eliminating the use of additional transformer components.[15] When employing such ferroresonant saturating-reactor circuits, as used in some types of frequency meters,[16] the average value of output voltage; i.e., the supply voltage of the magnetic amplifier, is linearily proportional to power-supply frequency: $E_P = \text{constant} \times f_P$. Thus, the flux density $B = \text{constant} \times E_P/f_P$ in the magnetic-amplifier cores will be substantially independent of voltage fluctuations of ±10 to 20%, and frequency variations of ±5 to 10%. In this way, by replacing the ordinary power-supply transformer of the magnetic amplifier by a small nickel-iron-alloy constant voltage transformer,[17] it is possible to eliminate the undesirable effects of voltage drift and frequency drift, and to obtain improved performance of low-level magnetic amplifiers.

Operation of self-balancing magnetic amplifiers with magnetic voltage stabilizers makes it possible to combine the specific advantages of both artifices and to obtain optimum performance. This is, of course, particularly significant for low-level d-c signal applications and for solutions of those special problems where extremely small values of linearity error and zero drift represent prime requirements.

MEASUREMENTS ON HALF-CYCLE-RESPONSE MAGNETIC AMPLIFIERS

The transient response of magnetic amplifiers with rectangular-hysteresis-loop core material can be described by simple linear equations.[18] When operating in the linear region of the output-input characteristic, the response of the load current to changes of d-c control voltage, power-supply voltage, and load resistance is an exponential function, governed by a "time constant," which varies with the load resistance and inversely with control-circuit resistance and power-supply frequency. It follows from equation 1 that the "control time constant" T_C can be expressed as follows:

$$T_C = \frac{1}{4f_P} K_P \frac{R_O}{R_L} \frac{1}{K_{AT}} \frac{1}{k_f^2} \text{ seconds} \qquad (3)$$

or

$$T_C = \frac{1}{4} K_P \frac{R_O}{R_L} \frac{1}{K_{AT}} \frac{1}{k_f^2} \text{ cycles} \qquad (4)$$

It is also possible to express the control time constant T_C by introducing the overall volts-per-turn gain. For example, the time-constant relationship of a self-saturating magnetic-amplifier circuit is

$$T_C = \frac{1}{2f_P} \frac{d(E_T/N_A)}{d(E_S/N_C)} \text{ seconds} \qquad (5)$$

or

$$T_C = \frac{1}{2} \frac{d(E_T/N_A)}{d(E_S/N_C)} \text{ cycles} \qquad (6)$$

where $E_T = E_L R_O/R_L$ is the average value of the total output voltage, N_A is the number of turns of the output (gate) windings, E_S is the average value of the input (control) voltage, and N_C is the number of turns of the input (control) windings.

In considering the accuracy of this theoretical method for determining the speed of response of conventional types of magnetic amplifiers having a response time of about 6 to 30 cycles, or more, it must be remembered that in most practical design problems an approximate value is all that is required. However, in regard to the transient performance of high-speed magnetic amplifiers having a response time of the order of 1 to 4 half-cycles of the power-supply frequency, experimental methods have assumed increasing importance in their development.

Convenient and accurate methods for analyzing the dynamic performance of slow-response magnetic amplifiers consist in applying half-cyclic integrators which supply information about the half-cyclic average of the output throughout a transient[19] or in the steady-state with periodic signals.[20] The dynamic performance of fast-response magnetic amplifiers may be analyzed by means of a transient analyzer[21,22] which produces a synchronous square-wave signal voltage and permits

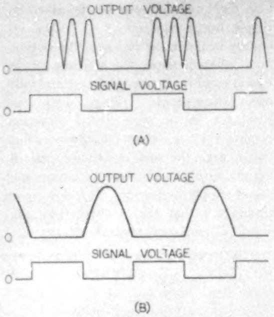


Fig. 1. Schematic waveshape diagrams characterizing half-cycle transient response of self-balancing magnetic amplifiers

A—The amplifier is 1,200-cycle-operated from a 400/1,200-cycle magnetic-frequency tripler, whereas the synchronous 400-cycle square-wave signal voltage is derived from a ferroresonant circuit

B—The amplifier is 400-cycle-operated directly from the power supply, and the synchronous 400-cycle square-wave signal voltage is derived from a ferroresonant circuit, which is also fed from this power supply

direct visual observation of a stationary pattern on a dual-beam oscilloscope. Application of the simple magnetic-amplifier transient analyzer employing a synchronously operated mechanical-switch device[21] is limited to low power-supply frequencies (e.g., 50 to 60 cps), where a mechanical switch can be operated in a reliable way. When working with higher power-supply frequencies (e.g., 400 to 2,000 cps), the synchronously operated switch device representing a mechanical modulator may be replaced by any suitable type of electric modulator; e.g., a magnetic switch[22] which consists essentially of a special saturable-reactor circuit producing an auxiliary current with rectangular waveshape. In this case, the magnetic amplifier under test may be supplied from a magnetic frequency multiplier[23] which produces an exact multiple of the frequency of the a-c power supply feeding directly the saturable-reactor circuit acting as a magnetic switch.

Fig. 1(A) shows schematically a typical waveshape diagram characterizing the half-cycle transient response of a single-stage self-balancing magnetic amplifier, as described previously.[24] The lower trace indicates the waveshape of the synchronous 400-cycle unidirectional square-wave signal voltage derived from the phase-shifter voltage of a magnetic-switch circuit. The upper trace shows the half-cycle transient response of the output voltage of this amplifier which is 1,200-cycle-operated from a 400/1,200-cycle magnetic frequency tripler.

A simplified and very convenient method for studying the transient performance of half-cycle response magnetic amplifiers consists of eliminating the frequency tripler, and feeding both the amplifier under test and the phase shifter of the magnetic-switch circuit directly from the a-c power supply; e.g., both with 60 or 400 cps. Fig. 1(B) shows schematically the respective waveshape diagram corresponding to the special case that the phase relationship between square-wave signal voltage (lower trace) and half-cycle pulses of the magnetic-amplifier output voltage (upper trace) is exactly 180 degrees.

Fig. 2 illustrates the simplified test circuit for measurements on half-cycle response magnetic amplifiers. The 60-cycle or 400-cycle 3-phase power supply feeds both the magnetic-amplifier circuit under test, as described in later sections of this paper, and a phase shifter which operates a ferroresonant magnetic-switch circuit (magnetic voltage stabilizer) consisting of a nonlinear reactor T_P with almost rectangular magnetic characteristic (saturating reactor with Orthonol tape core), a capacitor C_P, and a linear reactor L_P. In this special case, the square-wave voltage E_{SQ} (60 cps or 400 cps, respectively) is applied to a splitting circuit carrying the square-wave alternating current I_M (without d-c component). This splitting circuit consists of two half-wave rectifiers (silicon junction diodes) D_M', D_M'', with potentiometer resistor R_M' ("black" first-half-cycle pulses I_M') and fixed resistor R_M'' ("white" second-half-cycle pulses I_M''). The load voltage E_L and the variable square-wave signal voltage E_S are applied to the channels V_1 and V_2 of the dual-beam oscilloscope. When using this test circuit, the dual-beam oscilloscope displays input and output voltages of the magnetic amplifier under test simultaneously and continuously on common time base. Thus, it is possible to study the effect of various circuit-parameter changes on actual dynamic performance and to determine optimum design.

Self-Balancing Push-Pull Magnetic-Amplifier Circuits for Operation of Ink Recorders

General

In magnetic-amplifier-operated ink recorders, push-pull action is used to obtain output polarity that depends on the signal polarity (duodirectional output-input characteristic). Two equally rated saturable-reactor systems (the "plus-minus system" and the "minus-plus system") are back-to-back connected in such a way that either the voltages produced by them across the load, or the currents drawn by them through the load, are in opposition. Other significant advantages of push-pull types of magnetic-amplifier circuits reside in the facts that the output-input characteristic passes through the origin, that linearity and stability are considerably improved, and that multi-stage design (cascading of two or three stages) is greatly facilitated. These additional advantages are so important that push-pull arrangements are sometimes employed in those cases where duodirectional output-input characteristic is not required; i.e., when the signal voltage or signal current does not change its polarity.

Generally, four saturable-reactor elements are incorporated in a symmetrical a-c bridge or differential-type circuit operating on a-c bridge balance conditions: Output is zero if input is zero. In practice, it is required that fulfillment of this condition (the zero stability) must be mostly independent of changes in magnitude, frequency, and waveshape of the

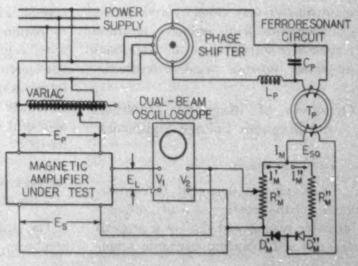


Fig. 2. Simplified test circuit for measurements on half-cycle-response magnetic amplifiers using the new method, as characterized by the schematic waveshape diagram of Fig. 1(B)

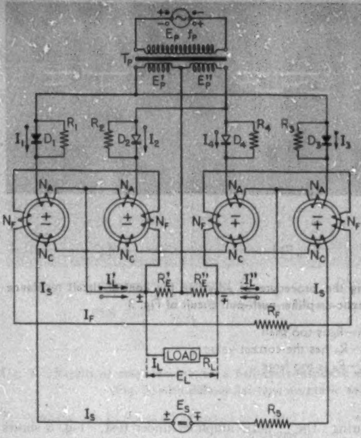


Fig. 3. The basic self-saturating push-pull circuit with differential-type magnetic negative feedback

power-supply voltage, major changes in ambient temperature, etc. In other words, the zero drift of the magnetic-amplifier circuit with regard to such changes in operating conditions shall be as small as possible. According to conventional practice in defining the performance of electric instruments, it is advisable to distinguish between drift-rate components, such as voltage drift, frequency drift, temperature drift, etc.

Feedback in Magnetic-Amplifier Push-Pull Circuits

In the orginal forms of magnetic-amplifier push-pull circuits with two single-ended external-feedback or internal-feedback (self-saturating) circuits in back-to-back connection,[25] each of these circuits has its own positive-feedback component which reduces or nearly nullifies the intrinsic negative feedback of the ordinary saturable reactor (transductor). These individual feedback components, however, are unaffected by the output-circuit design.

In contrast to these original forms of push-pull arrangements, magnetic amplifiers of the differential-feedback type[26,27] utilize feedback derived from the load circuit, which is common to both saturable-reactor systems. The resultant differential feedback, which may be positive or negative as required, is proportional to the actual difference between the individual load currents and is zero with no-signal conditions. As a result, the magnitude of individual quiescent currents (load currents with no-signal conditions) is unaffected by differential feedback, which may be derived from either output voltage (differential voltage feedback) or output current (differential current feedback).

Differential feedback may be produced by additional windings, which are inductively coupled with the control windings (magnetic feedback[26,27]). In this case, the differential-feedback circuit does not have any influence upon the effective input resistance of the control circuit; and the effective control-circuit impedance is determined solely by the actual total resistance of this circuit, comprising the resistance of the control windings and a series resistor (forcing resistor) which is conventionally provided to increase the speed of response of the magnetic amplifier. These arrangements are sometimes termed "self-excited" or "pseudo-feedback" circuits, because they do not have the inherent property of influencing the actual magnitude of the input-circuit resistance, as for example, the electronic amplifier of the cathode-follower type does.

In 1944, the author[28,29] showed that "true feedback" ("electric" feedback) can be applied to magnetic amplifiers, particularly for instrumentation purposes, by feeding back a voltage or current, proportional to the actual load current, resistance-coupled into the control-circuit loop so that a very large amount of degeneration is obtained. In this arrangement, a push-pull-type magnetic amplifier acting as a d-c/a-c converter or modulator is employed in conjunction with an ordinary a-c vacuum-tube amplifier and a phase-sensitive rectifier; and over-all negative differential feedback is utilized for stabilization of the system. Combination of this arrangement with a moving-coil ink recorder offers the possibility of making a permanent record of very small direct voltages or current created, for instance, by thermocouples, photoelectric cells, or radiation pyrometers.

This paper, however, is concerned only with all-magnetic devices employing no vacuum-tube or transistor components, and having inherent self-balancing potentiometer characteristics. Experience has shown that such magnetic-amplifier-operated ink recorders are preferred in many situations, particularly for long-time reliability testing of circuit components, where absolute reliability of the d-c amplifier, which operates the ink recorder, is of prime importance.

There are various types of magnetic-amplifier push-pull circuits utilizing either external feedback[30] or internal feedback (self-saturation[31]), which may be employed in conjunction with ink recorders. Self-balancing magnetic amplifiers of the differential-feedback type with bias excitation and external-feedback circuits have been described in detail previously.[32,33] These amplifiers contain a comparatively large number of rectifier elements (12 half-wave rectifiers in each stage). The simplified push-pull circuits described in the following sections, however, employ self-saturation and contain

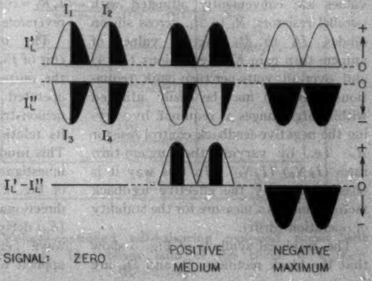


Fig. 4. Variations of the average values of the half-cycle pulses I_1 to I_4, and waveforms of the resultant full-wave unidirectional currents I_L', I_L'', and $I_L' - I_L''$ for the special case that, with no-signal conditions, the firing angle of the four saturable-reactor elements (Fig. 3) is 90 degrees

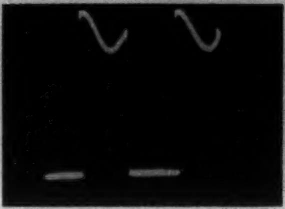


Fig. 5. Oscillogram of the waveform of the synchronous 400-cycle square-wave signal voltage, which is derived from a ferroresonant circuit, as shown in Fig. 2

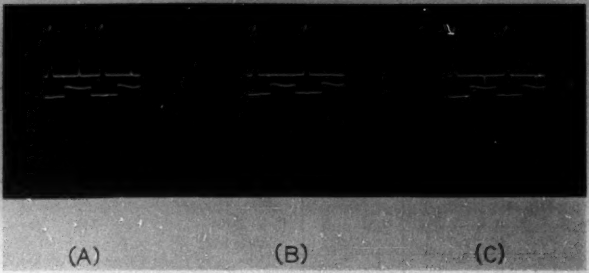


Fig. 6. Oscillograms characterizing the procedure of adjusting the control-circuit resistance R_S in the magnetic-amplifier-push-pull circuit of Fig. 3

A—R_S is too small
B—R_S has the correct value
C—R_S is too large

only four half-wave rectifier elements (preferably silicon junction diodes) in each stage. They have been used successfully in 2-stage and 3-stage designs. In these multistage arrangements, the input-stage circuit with Supermalloy tape cores and silicon rectifiers is supplied from a ferroresonant constant-voltage transformer to improve stability of operation with regard to zero drift and constant gain.

THE BASIC SELF-SATURATING PUSH-PULL CIRCUIT WITH DIFFERENTIAL-TYPE MAGNETIC NEGATIVE FEEDBACK

The basic self-saturating push-pull circuit[34] which may be employed either with a-c load or with d-c load, contains only four half-wave rectifier elements. In the typical example of Fig. 3, the two equally rated saturable-reactor systems (the plus-minus system and the minus-plus system) with gate windings N_A, d-c control windings N_C, and d-c differential-feedback windings N_F (providing magnetic negative feedback) are combined by means of a power-supply transformer T_P with a center-tapped secondary winding. The resultant voltage drop (load voltage) E_L across series-mixing resistors R_E', R_E'' is proportional to the actual resultant output current $(I_L'-I_L'')$, where $I_L'=I_1+I_2$, and $I_L''=I_3+I_4$. Quiescent-current values are conveniently adjusted with parallel resistors, R_1 to R_4, across silicon diodes, D_1 to D_4. Actual values of ampere-turn gain K_{AT} (equations 1 to 4) and over-all volts-per-turn gain (equations 5 and 6) may be easily adjusted within large ranges as required, by varying the negative-feedback control resistor R_F; i.e., by varying the ampere-turn ratio $(I_FN_F)/(I_SN_C)$. In this way it is possible to vary the effective feedback factor, which is a measure for the stability of operation (drift).

The graphical symbols of Fig. 3 show that the black rectifiers, D_1 and D_3, are conducting simultaneously during the first half-cycle period of power-supply voltage E_P (black polarities), while the white rectifiers, D_2 and D_4, conduct simultaneously during the second half-cycle period of E_P (white polarities). Fig. 4 shows the variations of the average values of the four half-cycle pulses, I_1 to I_4, and the waveforms of the resultant full-wave unidirectional output currents I_L', I_L'', and $I_L'-I_L''$ for the typical example that, with no-signal conditions $(E_S=0)$, the pulses I_1 to I_4 have 50% of their maximum value; i.e., the firing angle of the four saturable-reactor elements is 90 degrees.

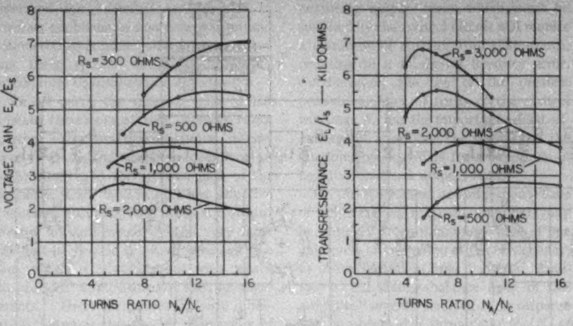
Evidently, with $E_S=0$, the average values of the first-half-cycle pulses, I_1 and I_3, and those of the second-half-cycle pulses, I_2 and I_4, are equal, and $I_L'-I_L''$ is zero. When a positive signal voltage, $+E_S$, is applied, the firing angles of I_1 and I_2 decrease, while those of I_3 and I_4 increase. However, when a negative signal voltage, $-E_S$, is applied, the firing angles of I_1 and I_2 increase, while those of I_3 and I_4 decrease. Thus, the resultant unidirectional full-wave output voltage $E_L=I_LR_L=\text{constant}\times(I_L'-I_L'')$ is polarity-reversible and is zero at zero input.

The magnetic-negative-feedback circuit of Fig. 3, which may be employed in the output stage of magnetic-amplifier-operated ink recorders, will serve as a demonstration of half-cycle response and its relationships to gain characteristics. This fundamental arrangement has been investigated using the simple test circuit of Fig. 2, in which a synchronous unidirectional square-wave signal voltage (E_S) derived from a ferroresonant square-wave generator with phase shifter is applied to the input circuit of the magnetic amplifier under test. Fig. 5 shows the waveform of the 400-cycle signal voltage $E_S=I_SR_S$ which is synchronous with the 400-cycle power-supply voltage E_P of the push-pull amplifier (Fig. 3).

When performing these measurements, the circuit parameters $(R_S, N_C, \text{and } R_F)$ and the angular position of the rotor of the phase shifter are adjusted always in such a way that the half-cycle-response output-voltage waveform, as schematically indicated by the upper trace of Fig. 1(B), is obtained. Fig. 6 illustrates this procedure in the typical case that the number of turns of control-winding units N_C and negative-feedback-winding units N_F are constant, whereas the series resistor R_S is varied (in the vicinity of correct adjustment) within certain limits: (1) Besides the first-half-cycle pulses (I_1 and I_3), there still are second-half-cycle pulses (I_2 and I_4) of greatly reduced amplitude, i.e., the resistance R_S is too small. (2) There are only first-half-cycle pulses, and the second-half-cycle pulses disappear (half-cycle response condition); i.e., the resistance R_S has the correct value. (3) Besides the first-half-cycle pulses (I_1 and I_3), there are smaller second-half-cycle pulses (I_2 and I_4) of negative polarity; i.e., the resistance R_S is too large.

Using this method for establishing half-cycle-response operating conditions of the push-pull circuit of Fig. 3, the parameters $R_S=E_S/I_S$, N_C, and R_F have been varied. The average values of signal voltage E_S, signal current I_S, and load voltage E_L are measured by means of high-precision moving-coil instruments.

Fig. 7 shows results of measurements on this circuit with Orthonol 2-mil tape cores (inside diameter = 1.0 inch; outside



(A) (B)

Fig. 7. Results of measurements on a half-cycle-response magnetic-amplifier push-pull circuit (Fig. 3) with differential-type magnetic negative feedback

A—Voltage gain E_L/E_S as a function of turns ratio N_A/N_C with control-circuit resistance R_S as a parameter

B—Transresistance E_L/I_S as a function of turns ratio N_A/N_C with control-circuit resistance R_S as a parameter

diameter = $1^3/_8$ inches; tape width = 1/4 inches), and with

$E_P=E_P'=E_P''=115$ volts ($f_P=400$ cps); $N_A=1{,}600$ turns;

$N_C=100$, 150, 200, 300, or, 400 turns; i.e., $N_A/N_C=16.0$, 10.7, 8.0, 5.3, or 4.0, respectively;

$N_F=50$ turns; $R_E'=R_E''=1{,}000$ ohms; $R_L=10{,}000$ ohms; $R_S=E_S/I_S=300\ldots3{,}000$ ohms.

There is a linear relationship between average values of E_S and E_L. The linearity error; i.e., the deviation of the actual output-input characteristic from the ideal linear characteristic $E_L=f(E_S)$ is of the order of $\pm1\%$ of full-scale deflection of the output voltmeter ($R_L=10{,}000$ ohms), which represents the load in the test circuit, where no additional load has been used.

Fig. 7(A) shows the usable voltage gain E_L/E_S as a function of turns ratio N_A/N_C with input resistance R_S as a parameter. This presentation makes it evident that, within certain limits (e.g., with $R_S=1{,}000$ ohms, and $N_A/N_C=5.3$ to 16.0), approximately the same voltage gain (about 3.3 to 3.8) can be obtained with different values of N_A/N_C by proper adjustment of the negative-feedback control resistor R_F. In order to obtain higher values of voltage gain E_L/E_S, the turns ratio N_A/N_C must be increased, and input resistance R_S must be reduced.

Fig. 7(B) shows the transresistance E_L/I_S as a function of N_A/N_C with R_S as a parameter. Within certain limits; e.g., with $R_S=1{,}000$ ohms, and $N_A/N_C=5.3$ to 16.0, approximately the same value of E_L/I_S (about 3.4 to 4.0) can be obtained by proper adjustment of the negative-feedback control resistor R_F. In order to obtain higher values of transresistance E_L/I_S, the turns ratio N_A/N_C must be reduced, and input resistance R_S must be increased.

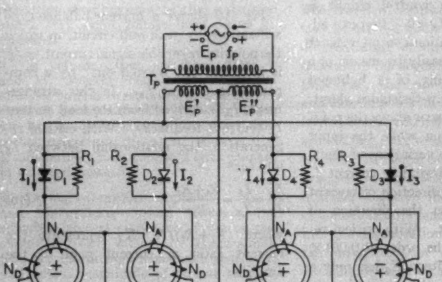


Fig. 8. Voltage-controlled self-balancing push-pull circuit with current feedback. The transconductance of this circuit is $I_L/E_S=1/R_K$

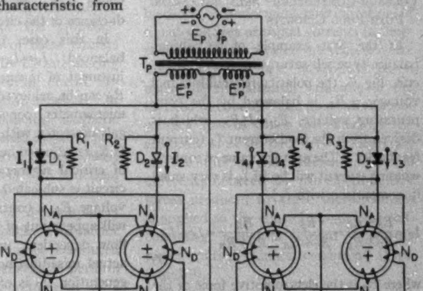


Fig. 9. Voltage-controlled self-balancing push-pull circuit with voltage feedback. The voltage ratio of this circuit is $E_L/E_S=(R_1+R_2)/R_1$

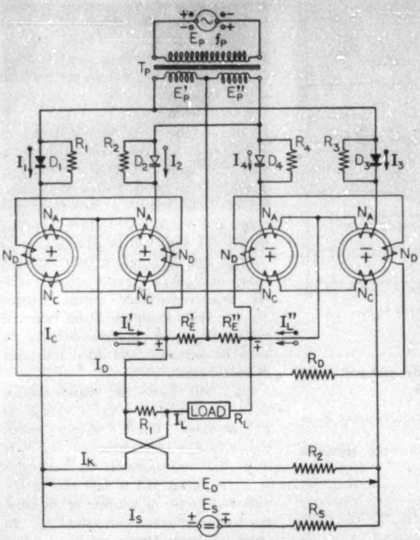


Fig. 10. Current-controlled self-balancing push-pull circuit with current feedback. The current ratio of this circuit is $I_L/I_S=(R_1+R_2)/R_1$

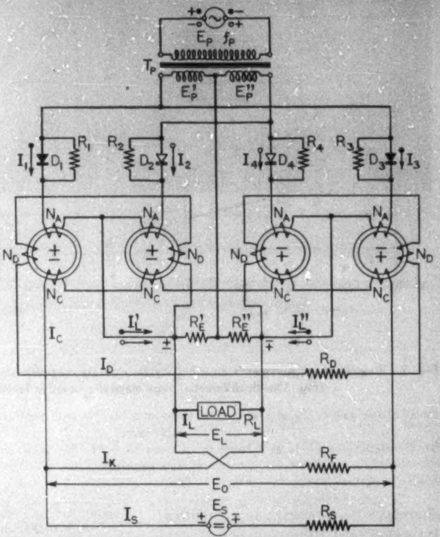


Fig. 11. Current-controlled self-balancing push-pull circuit with voltage feedback. The transresistance of this circuit is $E_L/I_S=R_F$

VOLTAGE-CONTROLLED SELF-BALANCING PUSH-PULL CIRCUITS

In the first example of a voltage-balance-type self-saturating push-pull circuit, Fig. 8, the polarity-reversible signal voltage, $\pm E_S$, is balanced by the compensating voltage, $E_K=I_LR_K$, which is derived from the load current I_L (current feedback). The average value of control-winding current will be (if I_C is very small in comparison with I_L)

$$I_C=\frac{E_S-E_K}{R_C+R_K}=\frac{E_S}{R_C+R_K}-\frac{E_K}{R_C+R_K}=I_S-I_K \tag{7}$$

where E_S is the electromotive force of the d-c signal source, and R_C is the total resistance of the control circuit between the terminals of negative-feedback resistor R_K (including the sum of the resistances of the control-winding units N_C, and the internal resistance of the signal source). The term $I_S=E_S/(R_C+R_K)$ is the signal-current component, and $I_K=E_K/(R_C+R_K)$ is the compensating-current componet. It has been shown[4] that with critical regeneration (infinite internal gain of the amplifier) the relationship between average value of load current I_L and signal voltage E_S is

$$\frac{E_S}{I_L}=R_K \quad\text{or}\quad \frac{I_L}{E_S}=\frac{1}{R_K} \tag{8}$$

where $1/R_K$ represents the transconductance of the circuit.

In this case, the control circuit is balanced: $I_C=I_S-I_K=0$. Proper adjustment of regeneration-control resistor R_D can be achieved easily by means of a microameter consisting of a lightspot galvanometer with low-resistance shunt: If R_D is adjusted below or above the point of critical regeneration while the input circuit is subjected to variations of signal voltage E_S, a control-winding current I_C will appear but of a direction of current flow dependent upon the direction of actual misadjustment. With critical regeneration, I_C is of the order of $0.001\times I_L$, and the signal voltage E_S has only a transient effect upon the magnetic amplifier.

When deriving the compensating voltage, E_K, from the load voltage (voltage feedback), the relationship between average value of load voltage E_L and signal voltage E_S is given by the resistance ratio of a voltage divider, as illustrated in Fig. 9:

$$\frac{E_L}{E_S}=\frac{R_1+R_2}{R_1} \tag{9}$$

where $(R_1+R_2)/R_1=K_E$ is the voltage ratio or voltage gain of the amplifier.

CURRENT-CONTROLLED SELF-BALANCING PUSH-PULL CIRCUITS

Fig. 10 shows a current-balance-type self-saturating push-pull circuit, in which the polarity-reversible signal current, $\pm I_S$, is balanced or "bucked out" by a compensating current I_K. In this arrangement, I_K is derived from the load current I_L (current feedback). With critical regeneration, the relationship between I_L and I_S is

$$\frac{I_L}{I_S}=\frac{I_L}{I_K}=\frac{R_1+R_2}{R_1} \tag{10}$$

where $(R_1+R_2)/R_1=K_I$ represents the current ratio or current gain of the amplifier.

In the modified arrangement of Fig. 11, the compensating current I_K is derived from the load voltage E_L (voltage feedback). The relationship between E_L and I_S is

$$\frac{E_L}{I_S}=\frac{E_L}{I_K}=R_F \tag{11}$$

where R_F is the transresistance of the circuit.

COMPARISON OF THE DIFFERENT FORMS OF SELF-BALANCING PUSH-PULL CIRCUITS

The adaptability of these circuits in conjunction with magnetic-amplifier-operated ink recorders with regard to various problems in instrumentation and telemetering may be judged by considering their specific technical properties which may be summarized as follows.

In all cases, the external gain of the system (transconductance, voltage ratio, current ratio, or transresistance, respectively) is substantially independent of large changes in the internal gain of the amplifier itself; and in all cases, effective resistance values of both input and output circuit are influenced by the very large amount of electric negative feedback in a way which is advantageous with regard to practical applications of such arrangements. There are, however, basic differences between the specific properties of the various forms of these self-balancing magnetic amplifiers, as illustrated in the typical examples of Figs. 8 to 11. These differences result from the different forms of negative-feedback networks and may be summarized in the following statements:

1. In voltage-balance amplifiers, Figs. 8 and 9, the effective input resistance E_S/I_C is greatly increased, and no appreciable current, I_C, is drawn from the signal source. These amplifiers act like a voltmeter having a very high resistance.
2. In current-balance amplifiers, Figs. 10 and 11, the effective input resistance E_O/I_S is greatly reduced, and no appreciable voltage drop, E_O, is introduced into the circuit under test. These amplifiers act like an ammeter having a very low resistance and a very small voltage drop across its terminals.
3. In voltage-output amplifiers, Figs. 9 and 11, utilizing voltage feedback, the effective internal resistance of the output circuit is greatly reduced. These amplifiers act like a low-impedance generator; and actual magnitude of output voltage, E_L, is substantially independent of large changes in load resistance R_L.
4. In current-output amplifiers, Figs. 8 and 10, utilizing current feedback, the effective internal resistance of the output circuit is greatly increased. These amplifiers act like a high-impedance generator; and actual magnitude of output current, I_L, is substantially independent of large changes in load resistance R_L.

It has been stressed already that the relationship between power gain, linearity error, and response time is such that if one property is accentuated, it is always at the expense of the other properties. Although the external gain of a self-balancing magnetic amplifier is independent of the number of turns of the control windings N_C, Figs. 8 to 11, it is generally desirable, in the interest of optimum linearity, to provide as many turns as the required response time will permit. In this way the inevitable residue of compensating ampere-turns which must be supplied by the control circuit will require a minimum of current from that source, resulting in a minimum linearity error. Furthermore, the use of the greatest possible number of turns of the control windings N_C has the important added advantage of improving the stability of the amplifier by providing for a maximum of negative-feedback ampere-turns.

On the other hand, actual response time of self-balancing magnetic amplifiers may be reduced to its minimum value of one half-cycle of the power-supply frequency, as characterized in Figs. 1 and 6. Analysis of the voltage-balance form of this amplifier[36] reveals that half-cycle response will be obtained if turns ratio N_A/N_C of self-saturating circuits (Figs. 8 to 11) is approximately equal to the resistance ratio R_O/R_K, where R_K is the actual resistance of the negative-feedback resistor, and R_O is the total resistance of the output circuit including the load-resistance components, copper resistance of the gate windings N_A, and forward resistance of the rectifiers, D_1 to D_4. Analysis of the current-balance form[36] of the self-balancing magnetic amplifier reveals similar relationships.

It is possible to apply nonlinear circuit elements; e.g., thermo converters, in the negative-feedback channel of the circuits of Figs. 8 to 11 to obtain special output-input characteristics of the amplifier. The transient response of these circuits may be altered materially by introducing differentiating or integrating circuit elements into the degeneration network. In this case, the self-balancing magnetic amplifier imposes its special dynamic performance characteristics on the associated system (e.g., ink recorder) with beneficial results. In accordance with common practice in the field of d-c amplifiers utilizing negative feedback, a capacitor or a mutual inductance may be introduced into the degeneration network to perform the integration or differentiation of a variable directly with respect to time.[37]

Algebraic Summation and Multiplication of Signals with Magnetic-Amplifier-Operated Ink Recorders

General

One of the more important advantages of magnetic amplifiers for instrumentation resides in the fact that complete isolation is possible between input and output. Such amplifiers are well suited to perform mixing of several d-c control signals at different voltage levels. Over the linear range of the output-input characteristic the output is proportional to the algebraic summation of signals. Thus, an obvious method to perform multiplication of signal voltages or signal currents consists in applying two thermocouples, one of which is heated by the algebraic sum of two derived currents and the other, by their difference. Insulated-heater-type thermo converters, however, are rather slow in response. Therefore, the use of such forms of thermocouples for this purpose is associated with an appreciable time lag (about 3 to 5 seconds) which may be permissible in some cases. When employing high-speed thermo converters,[38] e.g., special constructions of thermocouples, as used in high-speed thermal wattmeters,[39] a response time of the order of one-half second can be obtained.

L. A. Finzi and R. A. Mathias[40] have shown that it is possible to perform multiplication and other analytical operations by means of special magnetic-amplifier circuits, the components of which are single-core half-wave amplifiers of the kind described by R. A. Ramey.[41] The operation of these multiplying circuits, however, involves the basic assumption that each of the two signal voltages has a certain polarity which does not change. Otherwise, the circuits become inoperative.

In the following sections, two novel methods of performing multiplication of signal voltages or signal currents in magnetic-amplifier-operated ink recorders are described. The first method is based upon the fundamental principle of the thermal wattmeter[42] and uses two thermocouples, one of which is heated by the algebraic sum of two derived currents, and the other, by their difference. Actual difference between the two electromotive forces produced by these thermocouples is applied to the control circuit of a self-balancing push-pull circuit operating under voltage-balance conditions (Figs. 8 and 9).

The second method consists in replacing the two thermocouples by two special magnetic-amplifier multiplying circuits, which act as square-law means and provide the square term of the algebraic sum of two derived currents, and the square term of their difference, respectively. This method does not involve any assumptions on actual polarities of the two signal voltages or signal currents to be multiplied; i.e., each of the two d-c signals may change its polarity without any restrictions. This performance is one of the basic requirements of these instrumentation problems, which are concerned with multiplication of polarity-reversible d-c signal voltages or signal currents.

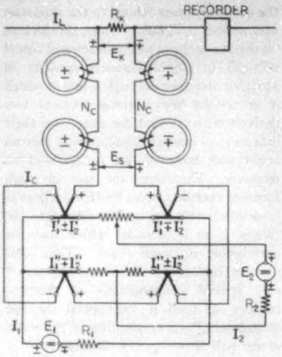


Fig. 12 (left). Multiplication of signal currents I_1 and I_2 by means of a symmetrical Wheatstone-bridge circuit with two pairs of insulated-heater-type thermo converters, and a self-balancing magnetic amplifier operating under voltage-balance conditions

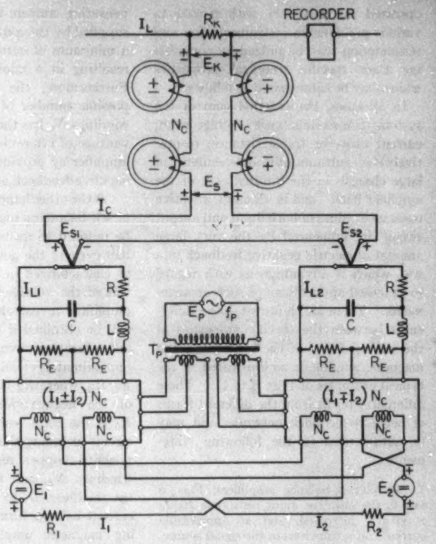


Fig. 13 (right). Multiplication of signal currents I_1 and I_2 by means of two directly heated thermal converters and two magnetic-amplifier push-pull circuits, which are controlled by (I_1+I_2) and (I_1-I_2) respectively

MULTIPLICATION OF SIGNALS WITH THERMOCOUPLES AND SELF-BALANCING MAGNETIC AMPLIFIER

Multiplication of signals with thermocouples is the fundamental principle of thermal wattmeters for measurement of power at frequencies above the conventional design range of electromagnetic instruments. In such arrangements, two thermocouples (or two groups of series-connected thermocouples) are employed, one of which is heated by the algebraic sum of two derived currents (I_1+I_2) and the other, by their difference (I_1-I_2). The quantity of heat produced in the two heaters of the thermal converters, assuming the heater resistance R_H to remain constant, is $R_H(I_1+I_2)^2$, and $R_H(I_1-I_2)^2$, respectively. With proper design, the temperature rise of the heated member is very closely proportional to the quantity of heat produced in it. Thus, when connecting the terminals of the two thermocouples in opposition, a d-c signal voltage, E_S, can be obtained, which is linearily proportional to the actual difference between the temperatures of the two heaters:

$$E_S=K'(I_1+I_2)^2-K''(I_1-I_2)^2 \qquad (12)$$

If the constants K' and K'' are equal, then

$$E_S=\text{const}\times[(I_1^2+2I_1I_2+I_2^2)-(I_1^2-2I_1I_2+I_2^2)] \qquad (13)$$

or

$$E_S=\text{const}\times 4I_1I_2 \qquad (14)$$

Because the square terms cancel in this case, the d-c signal voltage E_S, indicated by an ordinary moving-coil millivoltmeter, is linearily proportional to the product $I_1\times I_2$ to be measured.

It has been stressed previously[4,33] that voltage-balance-type self-balancing magnetic amplifiers are particularly well suited for measuring small d-c millivoltages created by thermocouples or radiation pyrometers. In the present application, this type of magnetic amplifier has been employed in conjunction with a standard-type moving-coil ink recorder for measuring and recording the d-c signal voltage $E_S=\text{const}\times I_1\times I_2$ (equation 14). Fig. 12 shows a symmetrical Wheatstone-bridge circuit with two pairs of insulated-heater-type thermo converters, one of which is heated by the algebraic sum of currents

$$(I_1'\mp I_2')=(I_1''\mp I_2'')=0.5(I_1\mp I_2) \qquad (15)$$

and the other, by the difference of these currents

$$(I_1'\pm I_2')=(I_1''\pm I_2'')=0.5(I_1\pm I_2) \qquad (16)$$

The individual currents $I_1=\text{const}\times E_1$ and $I_2=\text{const}\times E_2$ to be multiplied are led to equipotential terminals of the bridge network to secure proper distribution of current components, as defined by equations 15 and 16.

The resultant polarity-reversible d-c signal voltage E_S (equation 14) produced by the series-opposing-connected pairs of thermocouples (Fig. 12) is balanced by the voltage drop $E_K=I_LR_K$ across negative-feedback resistor R_K of the self-balancing magnetic amplifier of Fig. 8. Thus, there is a linear relationship between $E_S=E_K$ and the average value of polarity-reversible load current I_L flowing through the ink recorder:

$$I_L=\text{const}\times I_1\times I_2=\text{const}\times E_1\times E_2 \qquad (17)$$

It is to be noted, however, that insulated-heater-type thermo converters, as used in the bridge circuit of Fig. 12, are rather slow in response, actual response time being about 3 to 5 seconds. Although such slow-response performance may be acceptable in some cases, it will prevent proper recording of phenomena acting at a moderately rapid rate. Considerably faster response, of the order of 1 or 2 seconds, can be obtained by using directly-heated high-speed thermo converters, or a special construction, where the thermocouples function also as heaters.[39] With this arrangement, response time of the order of one-half second is obtained.

Application of directly heated thermo converters is associated with the basic requirement that the heaters of the two thermocouples are insulated from each other. The arrangement of Fig. 13 meets this requirement by providing two magnetic-amplifier push-pull circuits, the control-winding units N_C of which are insulated from the two output circuits being also insulated from each other (by means of two separate center-tapped secondary

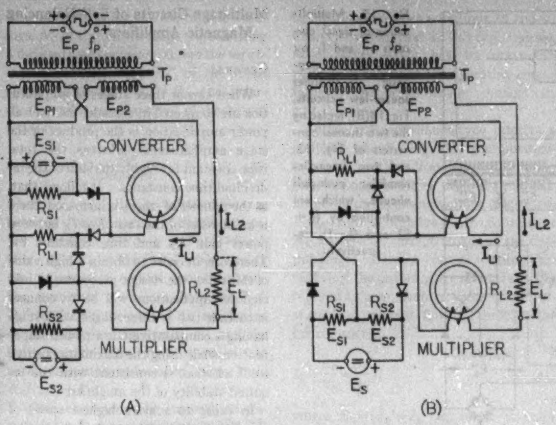


Fig. 14. Basic forms of magnetic-amplifier multiplying circuits

A—Original circuit which multiplies two d-c signal voltages E_{S1} and E_{S2} (Finzi and Mathias[40])
B—Modified circuit, which multiplies the components $E_{S1}=E_{S2}=0.5\times E_S$ of a single d-c signal voltage E_S, and may be used for providing square terms of signal voltages and signal currents (square-law-circuit output voltage $E_L=\text{constant}\times E_S^2$)

windings of a common power-supply transformer). Actual magnitudes of output currents I_{L1} and I_{L2} depend here on the total magnetomotive force of control produced jointly by the input windings N_C carrying currents $I_1=E_1/R_1$ and $I_2=E_2/R_2$, respectively. In this case,

$$I_{L1}=\text{const}\times(I_1\pm I_2)N_C \tag{18}$$

and

$$I_{L2}=\text{const}\times(I_1\mp I_2)N_C \tag{19}$$

where I_{L1} and I_{L2} are the average values of the two output currents. The two filter circuits are provided to secure that the rms values of heater currents are substantially equal to the average values of I_{L1} and I_{L2}. If this condition is fulfilled; i.e., if the a-c components of I_{L1} and I_{L2} are reduced to be only about 2 to of 3% the associated d-c components, then

$$E_{S1}=\text{const}\times I_{L1}^2=\text{const}\times(I_1\pm I_2)^2 \tag{20}$$

and

$$E_{S2}=\text{const}\times I_{L2}^2=\text{const}\times(I_1\mp I_2)^2 \tag{21}$$

The resulting d-c signal voltage

$$E_S=E_{S1}-E_{S2}=\text{const}\times I_1I_2 \tag{22}$$

is balanced by the voltage drop $E_K=I_LR_K$ across negative-feedback resistor R_K of a self-balancing push-pull circuit (Fig. 8). In this case, too, there is a linear relationship between $E_S=E_K$ and the average value of the polarity-reversible load current I_L flowing through the ink recorder:

$$I_L=\text{const}\times I_1\times I_2=\text{const}\times E_1\times E_2 \tag{23}$$

The accuracy of these thermal methods (Figs. 12 and 13) to perform multiplication of signals is such that the differences between the actually indicated values and those corresponding to an ideal performance characteristic (equation 23) is about ±2 to 3% of full-scale deflection of the ink recorder. This accuracy is considered to be sufficient for many applications. There is, of course, still a basic limitation which is caused by the relatively slow response of thermal converters (minimum response time about one-half of one second).

Multiplication with Magnetic Amplifiers and Magnetic-Amplifier Square-Law Circuits

In order to make operation of such multiplying devices in conjunction with an ink oscillograph (e.g., Brush recorder) possible, the two thermocouples of Fig. 13 have been replaced by two special magnetic-amplifier multiplying circuits, which act as square-law means and provide the square term of the algebraic sum of two derived currents, and the square term of their difference, respectively. In this arrangement, full advantage is taken of the half-cycle-response characteristics of Ramey-type single-core half-wave magnetic amplifiers, and full-wave push-pull circuits (Figs. 3 and 8) utilizing a very large amount of negative feedback.

The magnetic-amplifier multiplying circuit, Fig. 14(A), as described by Finzi and Mathias,[40] is basically a 2-stage arrangement, which contains the converter (input stage controlled by the first signal voltage E_{S1}), and the multiplier (output stage controlled by the second signal voltage E_{S2}). Resistors R_{S1} and R_{S2} provide the necessary path for the presetting magnetizing currents flowing against the signal voltages. Operation of the circuit of Fig. 14(A) can be described by the equation

$$E_L=I_{L2}R_{L2}=\text{const}\times\frac{E_{S1}\times E_{S2}}{E_{P1}} \tag{24}$$

where E_{S1} and E_{S2} represent average values, and E_{P1} is the rms value of supply voltage E_{P1} derived from the power-supply transformer T_P having two separate secondary windings. With $E_{P1}=$ constant, there is the simplified relationship

$$E_L=\text{const}\times E_{S1}\times E_{S2} \tag{25}$$

Proper operation of this multiplying circuit, however, involves certain requirements, which may be summarized as follows:

1. In each stage, the gate voltage (E_{P1} and E_{P2}, respectively) must satisfy the condition: $E_{P\max}<\omega N\Phi_{\text{sat}}$, where $\omega=2\pi f_P$ is the angular velocity of the a-c power supply (frequency f_P), N is the number of turns of the cores, and Φ_{sat} is the saturation flux density of the rectangular-hysteresis-loop core material.

2. Certain limitations with regard to the relationships between the instantaneous values of signal voltages E_{S1}, E_{S2}, and gate voltages E_{P1}, E_{P2} require that E_{P1} and E_{P2} must be square waves symmetrical about zero (as obtained from a d-c power source through a rotating commutator[40]) if E_{S1} and E_{S2} are pure d-c signal voltages (as derived from battery or filter circuits).

3. In both stages, the magnetizing currents must be minimized by proper design of the high-permeability cores. Since the magnetomotive force required by a core over intervals of increasing or decreasing fluxes is practically constant, a large part of it may be provided by an auxiliary winding fed from a synchronous compensating current[40] which is derived from the same power supply (E_P).

4. Each of the two d-c signal voltages, E_{S1} and E_{S2}, has a certain polarity (with regard to the half-wave rectifier elements series-connected with the resistors R_{S1} and R_{S2}), which does not change. Otherwise, the circuit becomes inoperative.

However, one of the basic requirements in many problems concerning multiplication of d-c signals is that each signal may change its polarity. In order to fulfill

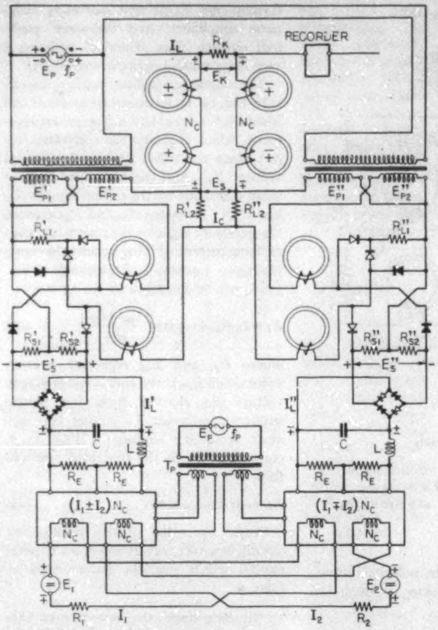


Fig. 15. Multiplication of signal currents I_1 and I_2 by means of two magnetic-amplifier square-law circuits, Fig. 14(B), replacing the two thermal converters of Fig. 13, and two magnetic-amplifier push-pull circuits, which are controlled by (I_1+I_2) and (I_1-I_2), respectively

this requirement, the two thermocouples in the arrangement of Fig. 13 have been replaced by two magnetic-amplifier square-law circuits, as shown in Fig. 14(B). The applied modification consists merely in supplying the two signal circuits from a center-tapped input resistor $(R_{S1}+R_{S2})$ so that $E_{S1}=E_{S2}=0.5\,E_S$. In this case, too, the signal voltage E_S must have a certain polarity, as indicated in Fig. 14(B). Otherwise, the circuit becomes inoperative.

In the complete arrangement of Fig. 15, the two magnetic-amplifier square-law circuits, Fig. 14(B), provide the square term of the algebraic sum of the currents $I_1=E_1/R_1$ and $I_2=E_2/R_2$, as well as the square term of the difference of these currents:

$$E_L'=\text{const}\times(I_L')^2=\text{const}\times(I_1\pm I_2)^2 \quad (26)$$

and

$$E_L''=\text{const}\times(I_L'')^2=\text{const}\times(I_1\mp I_2)^2 \quad (27)$$

Filter circuits L, C, and full-wave rectifiers across mixing resistors R_E (Fig. 3) are provided to convert the polarity-reversible output currents $I_L'=\text{const}\times(I_1+I_2)$ and $I_L''=\text{const}\times(I_1\mp I_2)$ of the two magnetic-amplifier push-pull circuits into substantially pure direct currents, which do not change their polarity.

The arrangement of Fig. 15 has been used in conjunction with a Brush recorder making a permanent record of the product of currents $I_1=0\pm50$ μa, and $I_2=0\pm50$ μa ($R_1=R_2=50{,}000$ ohms). When supplying the two magnetic-amplifier square-law circuits, Fig. 14(B), from a ferroresonant circuit or from a switching-transistor d-c to a-c converter ($f_P=400$ cps), the gate voltages (E_P) are symmetrical square waves symmetrical about zero; and pure d-c signal voltages E_S' and E_S'' can be applied in this case without disturbing proper operation of these circuits. Supermalloy 2-mil tape cores have been used successfully to minimize the magnetizing currents. Five silicon junction diodes in each circuit are connected, as shown in Figs. 14(B) and 15. The accuracy of measurements obtained with this arrangement (about ±2 and 3% of full-scale deflection of the ink oscillograph) was sufficient in the respective application, where high-speed operation (the response time is of the order of about 2 cycles of the power-supply frequency; i.e., about 5 msec) of the recorder is of major importance.

Multistage Circuits of Self-Balancing Magnetic Amplifiers

GENERAL

When two or three stages of amplification are connected in cascade, the over-all power amplification is the product of the stage amplifications, whereas the total time constant is roughly the sum of the individual time constants. It follows that, as the number of stages is increased, there is an increase of the ratio K_P/T_C between power gain K_P and time constant T_C. Therefore, in order to obtain a high value of this ratio, the design procedure in ink-recorder applications will be to connect in cascade two or three stages, each stage having a comparatively low power amplification while using the maximum positive feedback that is consistent with the required stability of the amplifier.

In order to achieve highest speed of response in multistage push-pull circuits, the parameters of each stage (Figs. 3, and 8 to 11) will be rated so that half-cycle response (Fig. 6) is obtained. When using several half-cycle-response full-wave push-pull circuits in cascade, the over-all response time is one half-cycle for first stage, and one half-cycle for each additional stage. Thus, the over-all response time of a 2-stage arrangement is one cycle and that of a 3-stage design is $1^1/_2$ cycles. In this way, with 400-cycle-operated multistage magnetic-amplifier circuits, an over-all response time of the order of 3 to 4 msec can be achieved.

INTERSTAGE COUPLING CIRCUITS

The various possibilities for connecting push-pull self-saturating circuits, Fig. 3, and Figs. 8 to 11, in cascade may be illustrated with regard to a 2-stage design, by three typical examples (Figs. 16 to 18). In the simplest form of interstage

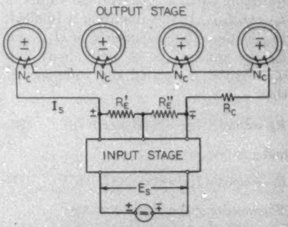


Fig. 16. Simplest form of interstage coupling circuit, in which the control windings N_C of the output stage are connected across the series-mixing resistors R_E', R_E'' of the input stage. The current ratio of this circuit is $I_S/(I_L'-I_L'')=R_E/(2R_E+R_C)$

coupling circuit, Fig. 16, the control windings, N_C of the output stage (representing a d-c load) are connected across the series-mixing resistors R_E', R_E'' of the input stage. With $R_E'=R_E''=R_E$, the polarity-reversible signal current of the output stage is

$$I_S=(I_L'-I_L'')\ \frac{R_E}{2R_E+R_C} \tag{28}$$

where $(I_L'-I_L'')$ is the actual difference between average values of resultant output currents $I_L'=I_1+I_2$ and $I_L''=I_3+I_4$ (Fig. 4). R_C is the total resistance of the output-stage control circuit between the terminals of mixing resistors R_E', R_E'' (including the sum of the resistances of the control-winding units N_C, and the series resistor, which may be omitted to obtain maximum value of resistance ratio $R_E/(2R_E+R_C)$.

When employing the current-balance-type interstage coupling circuit, Fig. 17, the control-winding current $I_C=I_S-I_K$ is substantially reduced to zero. Thus, actual voltage drop, E_O, across series-mixing resistors R_E', R_E'' is also substantially reduced to zero:

$$E_O=I_CR_N=(I_S-I_K)R_N=0 \tag{29}$$

In this case, the output-stage control circuit acts like an ammeter without voltage drop across its terminals; and the effective resistance of this control circuit is practically about 100 to 1,000 times smaller than actual resistance (R_N) of the four series-connected control-winding units (N_C). Consequently, with $E_O=0$,

$$I_S=0.5(I_L'-I_L'') \tag{30}$$

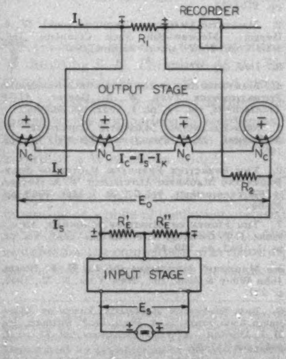


Fig. 17. Current-balance-type interstage coupling circuit, in which the control-winding current, $I_C=I_S-I_K$, of the output stage is substantially reduced to zero. The current ratio of this circuit is $I_S/(I_L'-I_L'')=0.5$

Comparison between equations 28 and 30 makes it evident that, in this case, current ratio $I_S/(I_L'-I_L'')$ has its maximum value with regard to the operating conditions of the series-resistor mixing circuit.

In the voltage-balance-type interstage coupling circuit, Fig. 18, the polarity-reversible voltage drop $E_L=f(E_S)$ across series-mixing resistors R_E', R_E'' is balanced by the opposing voltage drop $E_K=I_LR_K$ across negative-feedback resistor R_K in the load-current circuit of the output stage. Because control-winding current I_C of the output stage is substantially reduced to zero, average value of output-stage load current flowing through the ink recorder is

$$I_L=\frac{E_K}{R_K}=\frac{E_L}{R_K} \tag{31}$$

where $E_L=(I_L'-I_L'')R_E$ is the average value of output voltage of the input stage.

In all cases, Figs. 16 to 18, the input stage may operate as a self-balancing circuit under either voltage-balance (Figs. 8 and 9) or current-balance (Figs. 10 and 11) conditions. When employing multistage push-pull circuits to perform multiplication of signals, a magnetic-negative-feedback input-stage circuit with two separate groups of control windings (Figs. 13 and 15) will be used.

Applications and Performance Characteristics of Magnetic-Amplifier-Operated Ink Recorders

General

The high sensitivity of Esterline-Angus permanent-magnet moving-coil ink recorders permits in many cases direct operation of these instruments without the use of amplifiers. There are two basic sensitivities, 0–1 ma (resistance, 1,400 ohms), and 0–5 ma (resistance, 70 ohms), either type completing its swing across the $4^1/_2$-inch active chart width in about 1/2 second with critical-damping conditions.[43]

In order to extend the range of application of these instruments, a simple and reliable d-c amplifier will be required. With regard to the moderate response properties of such recorders, a 60-cycle-operated magnetic amplifier with a time constant of about 100 msec is generally acceptable. The arrangement is well suited for recording of various phenomena acting at a moderately rapid rate. However, actual requirements on power gain, linearity, and stability of operation indicate that a 2-stage design with self-balancing characteristics is needed for applications, where very good zero stability is essential.

By supplying the input stage of the 60-cycle 2-stage design with Supermalloy 2-mil tape cores and silicon rectifiers from a ferroresonant constant-voltage transformer,[33] excellent stability with regard to zero drift and constant gain has been obtained. When providing the signal-voltage range 0–5 mv or 0–±2.5 mv, this design is suitable for many applications with low-impedance signal sources; e.g., thermocouples, radiation pyrometers, strain-gage or resistance-thermometer bridge circuits, bolometer arrangements, and for resistance measurements by means of constant-voltage-supplied voltage-divider and bridge networks. With the signal-voltage range 0–20 mv or 0–±10 mv, half-cycle response is achieved. Zero stability is of the order of 100 µv per 8 hours or better at room temperatures. In a special application concerning long-time reliability testing of high-sensitivity relays, a 1-ma 1,400-ohm ink recorder in conjunction with a 60-cycle 2-stage magnetic amplifier has been operated as a high-impedance d-c millivoltmeter (8,000 ohms) having the range 0–20 mv or 0–2.5 µa for full-scale deflection of the recorder.

Brush-type ink oscillographs have been operated successfully in conjunction with a 400-cycle 2-stage or 3-stage magnetic amplifier. Dynamic performance of this arrangement can be tested conveniently by means of a motor-operated switch which closes and opens periodically an auxiliary battery-supplied test circuit, from which a unidirectional square-wave signal voltage is derived. The switch in this test circuit is operated so that, with

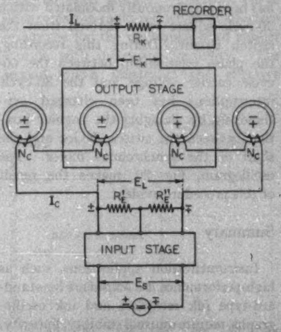


Fig. 18. Voltage-balance-type interstage coupling circuit, in which the control-winding current I_C of the output stage is substantially reduced to zero

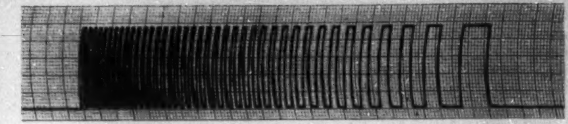


Fig. 19. Brush-recorder ink oscillogram of a 100-mv square-wave signal voltage in the frequency range from 30 cps down to d-c operating conditions. The attenuation does not exceed 5%

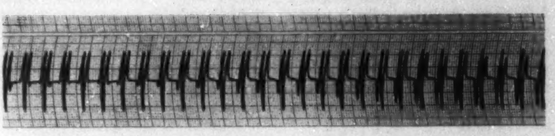


Fig. 20. Brush-recorder ink oscillogram, which has been obtained with a sinusoidal 60-cycle-carrier signal voltage being modulated with a 30-cycle square wave derived from a synchronous-motor-operated switch

full speed of the 60-cycle synchronous motor, a 30-cycle square-wave signal voltage is applied to the input circuit of the magnetic amplifier, the output current of which supplies the ink oscillograph. Then, after switching off the supply voltage of the motor, the frequency of the square-wave signal voltage is continuously reduced until the switch stops its action and remains in its open-circuit position. Fig. 19 presents an oscillogram of a 100-mv square-wave signal voltage, which has been recorded in this way.

The oscillogram of Fig. 20 shows a sinusoidal 60-cycle signal voltage, which is periodically interrupted by means of the synchronous-motor-operated switch in such a way that this voltage is turned on and off during 2 cycles, alternately. Thus, the 60-cycle-carrier signal voltage has been synchronously modulated with a 30-cycle square wave derived from the switch circuit. During this recording, the phase relationship between the 60-cycle carrier voltage and the 30-cycle interruptions has been changed continuously by changing the angular position between the switch device and the shaft of the synchronous motor. The oscillogram, Fig. 20, makes the result of this procedure evident.

Summary

Instrumentation applications, such as high-performance d-c amplifiers for standard-type ink recorders and ink oscillographs, require unusual stability, linearity, freedom from zero drift, and reliability which can be best met by an all-magnetic system using no vacuum tubes. A magnetic instrument amplifier which uses only nickel-iron-alloy tape core saturable reactor elements, silicon junction diodes, resistors, capacitors, linear reactors, and ordinary transformers can be built to meet the necessary performance requirements.

The inclusion of a multiplicity of voltage and current ranges, special control-winding arrangements, and auxiliary circuits makes magnetic-amplifier-operated ink recorders attractive for measuring signal voltages of the order of 10 to 100 mv, and signal currents of the order of 2 to 50 μa. This is possible without impairing accuracy and dynamic performance of such recorders.

Magnetic-amplifier-operated ink recorders are also well suited to perform mixing of signals; i.e., algebraic summation of signal voltages and signal currents. Multiplication of signals may be performed by means of two thermocouples, one of which is heated by the algebraic sum of two derived currents and the other, by their difference. Another possibility to perform multiplication consists in replacing the two thermocouples by two magnetic-amplifier multiplying circuits, which act as square-law means and provide the square term of the algebraic sum of two derived currents, and the square term of their difference, respectively.

Magnetic amplifiers of the self-balancing potentiometer type operating under either voltage-balance or current-balance conditions in conjunction with ink recorders offer important advantages with regard to power gain and dynamic performance. With proper design of these self-balancing magnetic amplifiers having infinite internal gain (critical regeneration) and utilizing a very large amount of electric negative feedback, resistance-coupled into the control-circuit loop, actual response time of the amplifier can be reduced to one half-cycle of the power-supply frequency. Thus, when using a 400-cycle power supply, the amplifier does reach 100% of its final steady-state output current 1.25 msec after the transient signal voltage is applied; and the output current drops to zero 1.25 msec after the signal voltage is suddenly removed.

A Brush recorder operating with a 400-cycle-operated 2-stage or 3-stage self-balancing magnetic amplifier (each stage having half-cycle response) has been used for recording 30-cycle signals with attenuation not exceeding 5%. The dynamic performance of this arrangement is characterized by typical ink oscillograms, which have been obtained by recording variable-frequency sinusoidal-waveshape and square-wave signals.

When employing a permanent-magnet moving-coil Esterline-Angus recorder, the response time of the 60-cycle-operated magnetic instrument amplifier of the order of 100 msec or more is permissible. This arrangement is well suited for recording of various phenomena acting at a moderately rapid rate. For long-time reliability testing of high-sensitivity relays, a 1-ma 1,500-ohm, ink recorder with a 60-cycle 2-stage self-balancing magnetic amplifier has been employed as a high-impedance d-c millivoltmeter (8,000 ohms) having the range 0–20 mv or 0–2.5 μa for full-scale deflection of the recorder.

References

1. Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art, J. G. Miles. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2104–23.

2. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957.

3. *Ibid.*, pp. 271–78.

4. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, 1952, pp. 383–95.

5. Self-Balancing Magnetic Amplifier, W. A. Geyger. *U. S. Patent no. 2,700,130*, Jan. 18, 1955.

6. The Effective Feedback Factor of Self-Balancing Magnetic Amplifiers, W. A. Geyger. *AIEE Transactions*, vol. 77, pt. I, Mar. 1958, pp. 66–69.

7. The Figure of Merit of Magnetic Amplifiers, J. T. Carleton, W. F. Horton. *Ibid.*, vol. 71, pt. I, 1952, pp. 239–45.

8. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, p. 212.

9. The Effective Feedback Ratio of Magnetic Amplifiers, L. A. Finzi, G. F. Pittman, Jr., H. L. Durand. *AIEE Transactions*, vol. 71, pt. I, 1952, pp. 157–64.

10. See reference 8, pp. 77–79.

11. A Simple D-C Type of Instrument Transformer Having Real Current-Transformer Properties (Ein einfacher Gleichstrom-Messwandler mit echten Stromwandler-Eigenschaften), W. Krämer. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 58, 1937, pp. 1309–13.

12. A New D-C Type of Instrument Transformer for the Measurement of High D-C Voltages (Ein neuer Gleichspannungs-Messwandler zur Messung hoher Gleichspannungen), W. Krämer. *Ibid.*, vol. 59, 1938, pp. 1295–98.

13. Fundamentals of Magnetic Amplifiers for the Techniques of Measurement and Control (Grundlagen der magnetischen Verstärker für die Mess- und Regeltechnik), W. A. Geyger. Wissenschaftliche Veröffentlichungen aus den Siemens-Werken, Berlin, Germany, vol. 19, 1940, pp. 226–69.

14. Magnetic Voltage Stabilizers for Instrumentation (Elektromagnetische Spannungs-Gleichhalter für Messzwecke), W. A. Geyger. *Siemens-Zeitschrift*, Berlin, Germany, vol. 15, no. 9, 1935, pp. 464–67.

15. Automatic Current and Voltage Stabilizers (Selbsttätige Strom- und Spannungsregler), W. A. Geyger. *Archiv für Technisches Messen*, Munich, Germany, section J 062–7, Nov. 1934, and section J 062–8, Dec., 1934.

16. A New Pointer-Instrument-Type Frequency Meter (Ein neuer Zeiger-Frequenzmesser), G. Keinath. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, vol. 37, no. 21, 1916, pp. 271–74, Fig. 1, 2.

17. A High-Sensitivity Magnetic Amplifier of the Balance Detector Type for the Techniques of Measurement and Control (Ein hochempfindlicher magnetischer Nullstrom-Verstärker für die Mess- und Regeltechnik), W. A. Geyger. *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, Berlin, Germany, vol. 21, 1943, pp. 47–57.

18. See reference 2, pp. 285–87, pp. 293–94.

19. A Transient Analyzer for Magnetic Amplifiers, E. J. Smith. *AIEE Transactions*, vol. 72, pt. I, 1953, pp. 461–65.

20. The Cyclic Integrator—A Device for Measuring the Frequency Response of Magnetic Amplifiers, T. Dunnegan, Jr., J. D. Harnden, Jr. *Ibid.*, vol. 73, pt. I, 1954, pp. 358–65.

21. Magnetic Amplifier Transient Analyzer, W. A. Geyger. *Electronics*, New York, N. Y., vol. 26, no. 7, 1953, pp. 189–91.

22. Magnetic-Switch Transient Analyzer W. A. Geyger. *Ibid.*, vol. 29, no. 1, 1956, pp. 150–51.

23. See reference 2, pp. 322–26.

24. Measurements on High-Speed Magnetic Servo Amplifiers with Two-Phase Motor Load, W. A. Geyger. *AIEE Transactions*, vol. 75, pt. I, 1956, pp. 242–48.

25. Apparatus for Measuring Direct Currents or Voltages, W. A. Geyger. *U.S. Patent no. 2,338,423*, Jan. 4, 1944.

26. See reference 17, Fig. 3.

27. Arrangement for Measurement of Small Direct-Current Quantities (Einrichtung zum Messen von kleinen Gleichstromgrössen), W. A. Geyger. *German Patent no. 702,082*, Jan. 30, 1941.

28. A Method for Ink-Recording of Small D-C Voltages (Ein Verfahren zum Aufzeichnen kleiner Gleichspannungen mit Tintenschrift), W. A. Geyger. *Doctoral Thesis*, University of Innsbruck, Sept. 1944.

29. Ink-Recording of Small D-C Voltages (Aufzeichnung kleiner Gleichspannungen mit Tintenschrift), W. A. Geyger. *Archiv der Elektrischen Übertragung*, Stuttgart, Germany, vol. 3, 1949, pp. 165–73.

30. See reference 2, pp. 148–61.

31. *Ibid.*, pp. 223–31.

32. Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type, W. A. Geyger. *AIEE Transactions*, vol. 76, pt. I, 1957, pp. 39–46.

33. Sixty-Cycle Low-Level Self-Balancing Magnetic Amplifier of the Differential-Feedback Type, W. A. Geyger. "Conference on Magnetic Amplifiers," *AIEE Special Publication T98*, Aug. 1957, pp. 187–204.

34. See reference 13, Figs. 12, 16, 17, 25.

35. The Operation of the Self-Balancing Magnetic Amplifier, A. D. Krall, E. T. Hooper. *AIEE Transactions*, vol. 75, pt. I, 1956, pp. 79–84.

36. The Operation of the Current-Type Self-Balancing Magnetic Amplifier, A. D. Krall. *Ibid.*, vol. 76, pt. I, 1957, pp. 380–84.

37. Analog Methods in Computation and Simulation (book), W. A. Soroka. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 75–79, pp. 82–86.

38. Vacuum Thermocouples. *Catalog no. 87LA*, American Thermo-Electric Company, Los Angeles, Calif., p. 7 ("High Speed Line," contact type).

39. A New High-Speed Thermal Wattmeter, J. H. Miller. *AIEE Transactions*, vol. 60, 1941, pp. 37–40.

40. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, L. A. Finzi, R. A. Mathias. *Ibid.*, vol. 72, pt. I, 1953, pp. 455–61.

41. The Single-Core Magnetic Amplifier as a Computer Element, R. A. Ramey. *Ibid.*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 442–46.

42. Die Technik Elektrischer Messgeräte (book), G. Keinath. R. Oldenbourg, Munich, Berlin, Germany, third edition, 1928, vol. II, pp., 81–82.

43. Esterline-Angus Recording Permanent-Magnet, Moving-Coil Type Direct-Current Milliammeters. *Catalog Section 42-56*, The Esterline-Angus Company, Inc., Indianapolis, Ind., 1956, pp. 1–8, Figs. 5 and 8.

A Precision, Guarded Resistance Measuring Facility

F. H. WYETH
MEMBER AIEE

J. B. HIGLEY
NONMEMBER AIEE

W. H. SHIRK, JR.
NONMEMBER AIEE

THE bridge principle was first introduced by S. H. Christie in 1833. Sir Charles Wheatstone in 1843 called attention to Christie's work and applied the bridge to the comparison of resistances.[1] Since that time, the network has been known as a Wheatstone bridge.

The mechanical designs of Wheatstone bridges have followed the practice of mounting on a plate of insulating material the ratio arm blocks (plug controlled), the switches for the comparison arm rheostat, the pairs of binding posts for the unknown resistance, battery source, and galvanometer detector, as well as keys for controlling battery and detector. The most favored top plate material has been hard rubber, although Bakelite and more recent materials with dimensional stability have been tried.

The insulating properties of solid dielectrics have been studied and papers presented over the past decades.[2–4]

The insulation resistance between metal parts, such as binding post studs mounted on an insulating plate, involves simultaneous effects of both surface and volume characteristics. Moisture absorption and moisture adsorption in combination with lack of surface cleanliness are factors which influence the rapid reduction of effective insulation resistance at high relative humidities.

The papers cited, covering insulating properties of solid dielectrics, give data for many materials, and tests under various atmospheric conditions. For convenient reference, included as Fig. 1 is a graph showing insulation resistance measured between two metal binding posts mounted on 1/2-inch-thick hard rubber with a center-to-center spacing of $1^{1}/_{4}$ inches and metal-to-metal distance of 3/4 inch at the surface. The hard rubber is about 10 years old and was in raw material open storage. The first curve illustrates the rapid reduction in effective resistance for the uncleaned hard rubber sample with increasing humidity. The second curve illustrates the effective

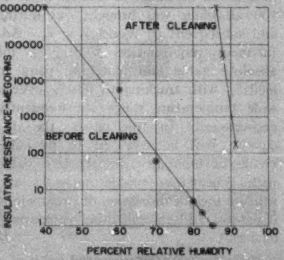
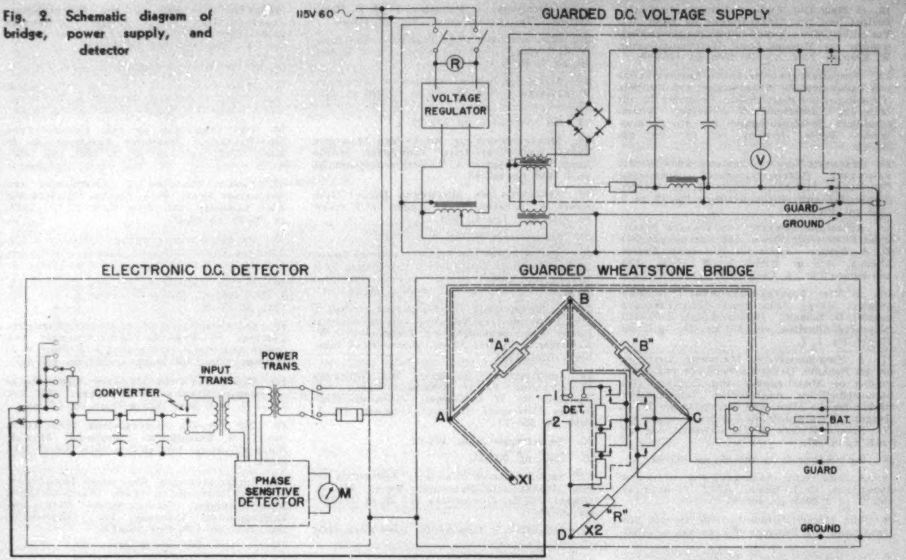


Fig. 1. Insulation resistance of hard-rubber top plate

Paper 58-556, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28–30, 1958. Manuscript submitted January 20, 1958; made available for printing March 4, 1958.

F. H. Wyeth, J. B. Higley, and W. H. Shirk, Jr., are with the Leeds & Northrup Company, Philadelphia, Pa.

Fig. 2. Schematic diagram of bridge, power supply, and detector



resistance after cleaning and subjecting it again to several humidities.

Two points can be made from these data.

1. The graphs shown represent possible shunt resistance in parallel with the unknown resistance across the X binding posts, but this may not be the entire story as there are other shunt paths between metal parts in the network which may contribute comparable errors.

2. The degree of cleanliness of the insulation and the amount of adverse humidity may not be known, and consequently no precise corrections can be applied to the measured value of the unknown resistance.

The need for precise adjustment of resistors in the range of 100,000 ohms to 100 megohms was brought out during and subsequent to World War II. Adjustment requirements to 0.01% for absolute value and matching pairs to 0.001% with tracking to 0.01% over a wide temperature range are frequently encountered. In this paper, the discussion will be confined to measurements of resistance for absolute value with Wheatstone bridge networks and will not include discussion of ratiometers, deviation bridges, and other comparison devices.[9]

If the error in the measured value of a nominal megohm resistor will be adversely affected by the effective insulation resistance to an amount dependent upon surface condition and humidity, it is evident that some method must be devised which will limit the error due to these uncontrollable environmental conditions to permit direct reading of the unknown absolute value to high precision.

A Guarded Wheatstone Bridge

The use of guarding to control the paths of direct currents as applied to the measurement of insulation resistance is a very old technique. Texts on electrical measurements have included circuitry with guarding to measure the insulation resistance of cables, and show guard rings to divert the current over the surface at the end of the cable back to the battery so that the current measured by the detector is only that passing through the insulation of the cable between conductor and sheath.[10]

It has not been customary to use guarding for d-c Wheatstone bridges.

Consider a Wheatstone bridge network, Fig. 2, in which the X_1 binding post, the ratio arms "A" and "B", and the connection from the ratio arm junction to the detector are enclosed in a guard circuit which is permanently connected to the C corner of the bridge. The guard is also carried back to the potential supply which will be described later.

In developing the apparatus described in this paper, a design requirement was that the insulation resistance between branch points in the assembled and wired bridge should exceed the following values:

A corner to guard, 10 megohms
B corner to guard, 10,000 megohms
C corner to ground (guard to ground) 10,000 megohms

The foregoing requirements apply for environmental conditions including up to 95% relative humidity. Note that leakage current from the A corner to guard is across the bridge and merely adds load to the power supply.

The solution was to use large plugs with blocks mounted on a hard rubber plate for the ratio arms. As part of the bridge manufacturing operation, the hard rubber was cleaned and then coated with 2% Dow Corning *Z920* silicone in chemically pure carbon tetrachloride. Tests on hard rubber, so treated, measured in excess of 10×10^{12} ohms between metal parts under 95% relative humidity. The ratio block panel is enclosed in the bridge with access through a door. The vertical position of the treated hard rubber panel and the enclosure should minimize dust and moisture collecting on the panel.

The hard rubber panel is supported at the four corners with Kel-F rods which have metal caps connected to guard.

These metal caps thus provide guarding of the entire ratio arm assembly, and the Kel-F support rods provide the high insulation resistance between guard and ground.

To summarize the bridge design up to this point, a guard system has been provided with specified limits for the insulation resistance requirements for the assembled bridge for the *A*, *B*, and *C* corners. It should be noted that the *C* corner to ground (guard to ground) insulation resistance becomes a shunt across the adjustable standard "*R*". As the maximum setting for this 6-decade standard is 111,111 ohms, the requirement of 10,000 megohms minimum for *C* corner to ground under all conditions will limit this shunting error to approximately 0.001%.

The bridge corners *A*, *B*, and *C* have been discussed, but the physical arrangement for the X_1 terminal which is connected to the *A* corner has not been described. Fig. 3 shows an exploded view and an assembly of a guarded binding post. The insulating material between the center conductor and the guard is acrylonitrile-styrene copolymer which is injection molded. The metal guard shell is installed in the bridge metal top plate using two similar plastic injection molded bushings.

Measurements made on binding post assemblies similar to that shown in Fig. 3 give insulation resistance values between conductor to guard and guard to ground exceeding 10×10^{12} ohms. As this plastic has exceedingly low moisture absorption, the insulation resistance exceeds 10^{12} ohms at 95% relative humidity.[11]

There are two features of this bridge network which should be mentioned. The first feature is that the large ratio arm resistors of woven manganin webbings have low time constants, matched temperature coefficient of resistance, and sufficient wattage rating to permit testing unknown resistances at maximum voltages as outlined in *MIL-R-93A* and *R-10509-B*. Maximum military specification voltages permit the use of lower sensitivity detectors.

The second feature is that screws are provided on the ratio arm blocks to connect directly to the bridge corners *A*, *B*, and *C*. This provides a means for checking the absolute values of the ratio arms for certification and maintenance.

Requirements for Guarded Bridge Power Supply

In the precision, guarded Wheatstone bridge described here, the d-c power supply is connected through a reversing switch to the *A* and *C* corners of the bridge. It is therefore necessary to eliminate the effects of insulation resistance between the *A* bridge-corner power-supply lead and ground, otherwise this resistance will shunt the unknown resistance being measured and may result in appreciable errors when making high resistance measurements. The first requirement, therefore, is that the power supply be guarded with the guard circuit connected to the *C* corner of the bridge.

As mentioned earlier, the insulation resistance between the *C* corner and ground represents a shunt across the standard arm. The second requirement of the power supply is that the guard to ground insulation resistance be at least 10,000 megohms.

A simple power supply meeting the foregoing requirements could consist of a battery resting on a metal plate which is insulated from ground by supports of a high-insulation-resistance material. The metal plate, being the guard, would be connected to the *C* corner of the bridge. It is also necessary that the battery leads to the bridge be air insulated, as a first choice. As an alternate, the leads may be contained in an insulated metallic braid sheath with the sheath connected to guard. The outer insulation for the sheath should be of a material such as polyethylene to maintain the high guard to ground insulation. With such a power supply, there is no possibility of a leakage current path directly from the *A* corner battery lead to ground. Insulation resistance between the *A* corner, battery lead and guard is, as mentioned previously, a shunt across the battery, and therefore need not meet rigid requirements.

For convenience in operation, a guarded-battery power supply should consist of multiple batteries and a switch for selecting the desired bridge potential. The voltage selector switch should be mounted on the guard plate or separately guarded.

A-C Operated Guarded D-C Bridge Supply

To eliminate the need for battery replacement or recharging, a guarded power supply operating from a 120-volt 60-cycle line would be more desirable for use with the precision, guarded Wheatstone bridge. The circuit is shown in the upper section of Fig. 2. To minimize the effects of line voltage variations, a voltage regulating transformer is used to supply the power to an adjustable autotransformer. The output of the autotransformer feeds the primary of the first isolation transformer which has its core and electrostatic shield connected to ground. The secondary of the first isolation transformer is connected to guard through the center-tapped primary of the second isolation transformer. Thus, these direct-connected transformer windings are the guard for the rectified output of the second isolation transformer.

The guard-to-ground insulation resistance is largely determined by the quality of the first isolation transformer since the insulation resistance of the secondary winding of this transformer is the controlling factor. Transformers meeting *MIL-T-27* specifications are required to have a minimum winding-insulation resistance of 10,000 megohms. Actual measurements on the isolation transformers used in this power supply showed insulation resistance in excess of 100,000 megohms at 90% relative humidity.

The rectifier and filter section are insulated from the guard circuit by the second isolation transformer. This transformer, the rectifier, and the filter components are physically mounted on an insulated metal guard plate. The voltmeter is mounted on the front panel but

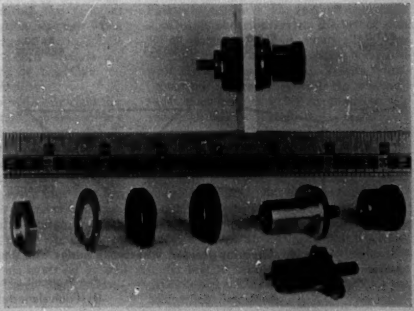


Fig. 3. Guarded binding post assembly

is separately guarded and insulated from the panel.

Although not shown in Fig. 2, a range-change switch is also incorporated in the guarded power supply with the switch guarded and insulated from the front panel. This switch changes the voltmeter range and voltage taps on the secondary of the second isolating transformer.

An *L-C* filter is used to provide a low internal impedance and therefore better regulation for various bridge settings. The output voltage is adjusted by setting the autotransformer, which has approximately full rotation for either output voltage range. These ranges are 0 to 10 and 0 to 100 volts d-c with up to 40 milliamperes output for any output voltage setting. The a-c component of the output voltage is less than 10 millivolts for any output current.

In operation, it has been found that the electrostatic shielding of the first isolation transformer is not perfect and results in the guard circuit having some a-c potential with respect to ground. This causes no trouble when using a galvanometer detector which is inherently a good 60-cycle mechanical filter. With a well-filtered electronic detector, there is no difficulty, but with an unfiltered electronic detector, a zero shift will be encountered. This undesirable effect can be eliminated by adding a filter network between the bridge and the unfiltered electronic detector.

Guarded Electronic D-C Null Detector

The desirability of measuring resistors accurately in the megohm range introduces the requirement of obtaining a detector with adequate sensitivity. This requirement has been met in the design of a stable, sensitive, low-noise detector with a filtered input and an acceptably high-input impedance. The detector is designed with a specially insulated floating-input circuit so that it may be used with one terminal grounded or both terminals ungrounded. The indicating element is a zero-center meter. The circuit is so designed that maximum sensitivity is obtained around the zero point and yet a nonlinear response characteristic allows extreme unbalance without off-scale deflections. The detector is also equipped with a sensitivity control switch and a zero adjust control. This latter control is necessary to adjust for the slight zero drift experienced during the initial warm-up period.

As presently designed, the detector has a 40,000-ohm input resistance and operates across approximately 100,000 ohms maximum in the bridge network. The sensitivity is sufficient under these conditions to show a 0.01% change in a 100-megohm resistor with 100 volts applied to the bridge. This condition represents approximately 0.3-microvolt change and corresponds to a 2-millimeter deflection. The maximum noise level and maximum zero drift after warm-up total less than one-half of this value.

A Method for Determining the Accuracy of High-Megohm Resistors Using Low-Resistance Standards

The requirement to construct d-c Wheatstone bridges to measure with high accuracies above the 1-megohm range has introduced the additional problem of checking these bridges for over-all performance under actual operating conditions. For values up to 1 megohm, the common practice has been to make absolute resistance measurements of the ratio and standard arms of the bridge and then to check these measurements by using the equipment to measure certified primary resistance standards. The ratio and standard arms of a Wheatstone bridge rarely exceed a value of 100,000 ohms and resistors up to these values can be checked to within 0.005% of their absolute value by comparison with certified standards. If a bridge is constructed to make measurements above the megohm range, the absolute value of the three arms of the bridge may be determined with a high degree of accuracy, but the final bridge check begins to lose its significance due to the lack of a high-accuracy certified primary standard.

A method may be used to simulate accurately primary standards above this range and provide a satisfactory over-all bridge check. The principle involved is that of the Y-delta transformation.[12] It is commonly known that impedances arranged in the form of a Y may be converted to another set of impedances in the form of a delta without affecting the values of the impedances between the three end points of the network.

Referring to Fig. 4, it can be seen that the Y circuit consisting of resistors R_{AO}, R_{CO}, and R_{DO} can be converted to the equivalent delta circuit R_{AC}, R_{AD}, and R_{CD} by means of the familiar transformations:

$$R_{AD}=R_{AO}+R_{DO}\quad\frac{R_{AO}\times R_{DO}}{R_{CO}}$$

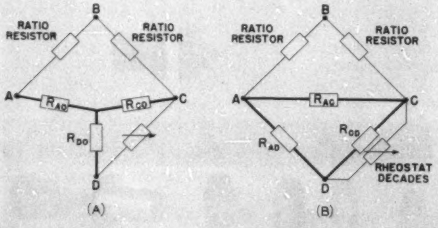
and

$$R_{AC}=R_{AO}+R_{CO}+\frac{R_{AO}\times R_{CO}}{R_{DO}}$$

and

$$R_{CD}=R_{CO}+R_{DO}+\frac{R_{CO}\times R_{DO}}{R_{AO}}$$

If R_{AO} and R_{DO} are each nominally 1-megohm resistors known to within 0.01% of their absolute value, and R_{CO} is a resistor whose value is determined from the table in Fig. 4 and is known to within 0.005% of its absolute value, then R_{AD} may be determined to a value within ±0.025%. The resulting value of R_{AC}



NOMINAL VALUE OF R_{AD}	R_{AO} (OHMS)	R_{CO} (OHMS)	R_{DO} (OHMS)	EQUIVALENT R_{AC} (OHMS)	EQUIVALENT R_{CD} (OHMS)
12 MEGOHMS	10^6	10^5	10^6	1.2×10^6	1.2×10^6
102 MEGOHMS	10^6	10^4	10^6	1.02×10^6	1.02×10^6
1002 MEGOHMS	10^6	10^3	10^6	1.002×10^6	1.002×10^6
10,002 MEGOHMS	10^6	10^2	10^6	1.0002×10^6	1.0002×10^6

Fig. 4. Checking guarded Wheatstone bridge with simulated standard megohm values using Y-delta transformation

A—Y circuit using certified standards
B—Equivalent delta circuit

Fig. 5. Precision, guarded resistance measuring facility

has no significant effect on the bridge balance. The value of R_{CD} does affect the bridge balance, but this effect may be precisely determined mathematically.

The simulated high megohm standards formed by this conversion permit the use of wire-wound resistance standards which have the advantages of high stability, low temperature coefficient, and low thermal properties. Precautions, however, must be observed in using these standards. The standards must be physically located in such a manner that the insulation resistance between the *A* and *D* points of the Y circuit shall be limited to an acceptable minimum. If the insulation resistance is low, the effect on the accuracy of the measurement will be the same as if this resistance were shunting the high megohm resistor formed by the Y-delta transformation. Proper use of a guarded terminal as shown in Fig. 3 at the *A* point of the Y network will minimize the errors due to this low insulation resistance.

This method of an over-all operational check on a Wheatstone bridge assembly provides an accurate means for checking the bridge assembly errors for the range of unknowns above 10 megohms. It also provides a practical method for checking bridge performance under ambient conditions of humidity.

Summary

The complete precision, guarded resistance, measuring facility as proposed is shown in Fig. 5. Each of the three components has been built; in fact, several of the precision, guarded bridges are in service. The complete guarding of the entire facility will insure a higher precision of measurement, lower maintenance, and provide a facility which can be used either in a modern standards laboratory or in non-air-conditioned production areas.

References

1. Electrical Measurements (book), **Forest K. Harris.** John Wiley & Sons, Inc., New York, N. Y., 1952.

2. Insulating Properties of Solid Dielectrics, **H. L. Curtis.** *Technical Bulletin*, National Bureau of Standards, *Scientific Paper no. 234*, vol. 11, 1915, pp. 369–417.

3. How Humidity Affects Insulation, Part 1 D-C Phenomena, **R. F. Field.** *General Radio Experimenter*, Cambridge, Mass., vol. 20, nos. 2 and 3, July-Aug., 1945.

4. A Method for Evaluating the Surface and Volume Resistance Characteristics of Solid Dielectric Materials, **W. G. Amey, F. Hamburger, Jr.** *Proceedings*, American Society for Testing Materials, Philadelphia, Pa., vol. 49, 1949.

5. Measuring of Surface and Volume Resistances, **R. K. Witt, J. J. Chapman, B. L. Raskin.** *Modern Plastics*, New York, N. Y., vol. 24, no. 8, Apr. 1947, p. 152.

6. The Formation of Ionized Water Films on Dielectrics under Conditions of High Humidity, **R. F. Field.** *The Journal of Applied Physics*, New York, N. Y., vol. 17, no. 5, May 1946.

7. Electrolytic Corrosion-Methods of Evaluating Materials Used in Tropical Service, **B. H. Thompson, K. N. Mathes.** *AIEE Transactions*, vol. 64, June 1945, pp. 295–99.

8. Sorption of Water by Plastics, **G. M. Kline, A. R. Martin, W. A. Crouse.** *Proceedings*, American Society for Testing Materials, vol. 40, 1940, pp. 1273–82.

9. Measuring Megohms to a Few Parts in a Million, **H. T. Wilhelm.** *Bell Laboratories Record*, New York, N. Y., vol. 25, Apr. 1947, p. 155.

10. Electrical Measurements (book), **F. A. Laws.** McGraw-Hill Book Company, Inc., New York, N. Y.

11. Modern Plastics Encyclopedia. New York, N. Y., "Plastics Properties Chart," 1957.

12. The Equivalence of Triangles and Three-Pointed Stars in Conducting Networks, **A. E. Kennelly.** *Electrical World*, New York, N. Y., vol. 34, 1899, pp. 413–14.

◆

Discussion

H. T. Wilhelm (Bell Telephone Laboratories, Murray Hill, N.J.): This paper makes several contributions to the art of precise d-c resistance measurement, particularly in the field of measurements in the megohm range.

Since 1943, shielded d-c bridges have been used in the Bell System to measure with a precision of a few parts per million the d-c resistance ratio of two resistors in the megohm range. These bridges are described in the April 1947 *Bell Laboratories Record*, cited as reference 9 of the paper. The shielding shown as a dashed-line rectangle in Fig. 2 of the afore-mentioned article is actually electrostatic shielding connected to ground, which serves two purposes: (1) to eliminate interference from changes in body capacitance and body static charges in the operator of the bridge, and (2) to avoid shunting the resistors under test with insulation leakage when measuring direct resistance ratio. This second objective agrees with what is accomplished by guarding in the Guarded Resistance Measuring Facility.

In the absence of formal definitions for guarding and electrostatic shielding, my idea of the difference between these terms is that guarding is essentially a technique for eliminating direct conductance between two electrodes, whereas shielding is a technique for eliminating direct capacitance between two electrodes. If this distinction is correct, it appears that this bridge and the bridge described by Mr. Wyeth are both guarded and shielded. It would be helpful to have formal definitions covering guarding, electrostatic shielding, and electromagnetic shielding.

N. E. Bonn (James G. Biddle Company, Philadelphia, Pa.): The authors have made a valuable contribution to the art of precision electrical measurement. The very effective use of the guard system combined with the application of the Y-delta transformation makes it possible to evaluate resistors as high as 10,000 megohms with a very high order of accuracy.

At the James G. Biddle Company, where the measurement of high resistance is an important part of the business, we are naturally interested in the design of such guarded bridges. But because our working standards are often subjected to high voltage, sometimes as high as 10,000 volts, we prefer a bridge arrangement which permits the application of such voltage to the resistor under test to take care of possible voltage dependency. This cannot be done with the bridge here described.

Of particular interest is the method of checking the performance of the bridge for multimegohm measurement by using available standards of relatively low resistance. This interesting, and presumably new, method deserves a more detailed discussion. Also, the data given in the table of Fig. 4 appear inadequate. It would probably have been helpful to those who are not expert in this field if the paper gave the procedure to be followed by the user in applying the Y-delta transformation, perhaps something like this:

1. Measure the unknown resistance X with the bridge in the regular manner to obtain a tentative value for $X=R_{AD}$. Record the reading R of the rheostat arm.
2. Calculate the value of R_{CO} from the first transformation equation using the tentative value of R_{AD} and the actual (not nominal) values of the standards R_{AO} and R_{DO}.
3. Arrange the Y circuit as in Fig. 4 using the two standard resistors and an accurate resistance box set to the calculated value of R_{CO}. Balance by adjusting rheostat arm and call this reading R'.
4. Calculate the value of R_{CD} using the third transformation equation.

The value of X obtained in the first measurement is correct only if

$$R=\frac{R'\times R_{CD}}{R'+R_{CD}}$$

If R_{AO} and R_{DO} are exactly 1 megohm each, as was assumed in the table of Fig. 4, the first and third transformation equations are easily simplified to read:

$$R_{CO}=\frac{10^{12}}{R_{AD}-2\times10^6}$$

and

$$R_{CD}=2R_{CO}+10^6$$

In this form the equations are certainly more comprehensive than the table since they may be applied to any value of X greater than 2 megohms.

It is interesting to note that with two 1-megohm resistors proposed in the paper, this check cannot be performed if X is equal to or smaller than 2 megohms, because R_{CO} either becomes infinity or assumes a negative value which means that the Y arrangement cannot be set up. If, however, instead of 1 megohm, R_{AO} and R_{DO} are each 1/2 megohm, the foregoing expressions for R_{CO} and R_{CD} become

$$R_{CO}=\frac{25\times10^{10}}{R_{AD}-10^6}$$

and

$$R_{CD}=2R_{CO}+5\times10^5$$

and can be used for checking the performance of the bridge for all values of X greater than 1 megohm. A 1-megohm standard with a tap in the middle would meet all requirements and would certainly be less costly than two 1-megohm standards of equal quality.

Additional references[1-3] on the use of guarding in d-c Wheatstone bridge circuits may be of interest.

REFERENCES

1. RECOMMENDED GUIDE FOR MAKING DIELECTRIC MEASUREMENTS IN THE FIELD. *AIEE Publication no. 62*, Apr. 1958.

2. MASTER TEST CODE FOR RESISTANCE MEASUREMENT. *AIEE Standard no. 550*, May 1949.

3. TENTATIVE METHOD OF TEST FOR ELECTRICAL RESISTANCE OF INSULATING MATERIALS. *ASTM Standard D257-57T*, American Society for Testing Materials, Philadelphia, Pa.

F. H. Wyeth, J. B. Higley, and **W. H. Shirk, Jr.:** We are grateful to H. T. Wilhelm for bringing out in his discussion the lack of formal definitions for "guarding" and "shielding." We are suggesting to the AIEE Committee on Terminology that consideration be given to including these terms in the ASA-AIEE definitions.

The Guarded Resistance Measuring Facility on view during the meeting had been subjected to 86% relative humidity and 85 degrees Fahrenheit for over 24 hours in an all-weather test chamber. The externally located certified standards were connected to the equipment with shielded leads. The equipment was operated under the high-humidity condition through ports in the chamber. The results confirmed the theory and practical effectiveness of the guarding.

We also appreciate the comments of Mr. Bonn. The resistance measuring facility described in this paper is designed for measurements of resistors at potentials up to 100 volts, which is the maximum permitted under military specifications such as *MIL-R-93A*, *R-10509-B*, and *R-14293A*. Lower potentials are permissible, providing the detector has adequate sensitivity. We recognize that there may be a few special applications where higher voltages on the unknown resistors would be required and Mr. Bonn has mentioned one such application. In a precision Wheatstone bridge, however, it is not economically practical to make provisions for such high testing potentials as required by Mr. Bonn.

The comments by Mr. Bonn in connection with the Y-delta transformation are mathematically correct and illustrate a procedure for checking the bridge at any value of unknown resistance above 1 megohm. The suggested procedure, however, requires the use of an adjustable resistance standard. We wish to point out that to maintain a ±0.025% limit of error as stated in the paper, this adjustable standard must be known to 0.005% and introduce no leakage paths. These are difficult requirements to meet. It is our belief that fixed check points using combinations of fixed standards as given in the table of Fig. 4 are satisfactory. Other combinations may, of course, be used. It should be remembered that the individual bridge resistors and ratios can be checked for accuracy. The simulated high megohm standards are, therefore, used as an over-all check, including the effectiveness of the guarding and insulation.

Vol.
Pt. 1
1958

A-C Controlled Magnetic Amplifiers

E. W. LEHTONEN
ASSOCIATE MEMBER AIEE

E. A. CRONAUER
ASSOCIATE MEMBER AIEE

PAST PRACTICE in a-c controlled magnetic amplifiers has required a relatively large resistance in series with the control windings to limit the circulating currents caused by the induced emf (electromotive forces) of the control windings. This resistor necessarily wastes signal power, and the realizable gain is much less than could be obtained from a d-c controlled amplifier.

To achieve higher gain with a-c control, an alternate method has been to demodulate the signal before using it to control the amplifier. This has several disadvantages, including drift, loss of signal power, and extra components.

This paper describes a method for controlling full-wave amplifiers with a-c signals which offers the following characteristics: no current-limiting resistors, no demodulator, response similar to d-c control amplifiers, and high gain. These characteristics are realized by arranging the control circuit so that the induced emf in the control circuit are substantially cancelled.

Among the other advantages of this type of amplifier are: the power efficiency of conventional full-wave amplifiers can be realized; and the high power gain makes it possible to control a relatively powerful magnetic amplifier with a low-power class *A* transistor, permitting the advantages of transistors and magnetic amplifiers to be combined.

Paper 58-773, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 1, 1958.

E. W. LEHTONEN and E. A. CRONAUER are with the Sperry Gyroscope Company, New Hyde Park, N. Y.

The authors acknowledge the assistance of L. Demuth in the measurements included in this paper.

Steady-State Description

Assume the full-wave amplifier circuit of Fig. 1, and assume further that by unspecified means the firing angle α is set somewhere between zero and π. The cores are assumed to have the hysteresis loops shown in Fig. 2 with the flux level indicated at the instant the sinusoidal excitation voltage is going through zero to positive values. If voltage drops caused by the current i indicated in Fig. 1 are temporarily ignored during the half-line cycle when $0<\omega t<\pi$, e_{ca}, e_{R_1}, e_{R_2}, and e_t will have the values shown.

To find the voltage across N_c of core B,

$$e_{cb}=-e_{ca}-e_{R_1}+e_{R_2}+e_t \tag{1}$$

The voltage across R_2 is zero since CR_2 is blocking.

$$e_{ca}=\frac{N_c}{N_p}E\sin\omega t\Big]_0^{\alpha}$$

$$e_{R_1}=\frac{R_1}{R_1+R_L}E\sin\omega t\Big]_{\alpha}^{\pi}$$

Substituting these values for e_{ca}, e_{R_1}, and e_{R_2} yields

$$e_{cb}=-\frac{N_c}{N_p}E\sin\omega t\Big]_0^{\alpha}-\frac{R_1}{R_1+R_L}E\sin\omega t\Big]_{\alpha}^{\pi}+0+E_t\sin\omega t \tag{2}$$

By adjusting the turns ratio of the transformer so that

$$E_t\sin\omega t=2\frac{N_c}{N_p}E\sin\omega t \tag{3}$$

choosing R_1 so that

$$\frac{R_1}{R_1+R_L}E\sin\omega t\Big]_{\alpha}^{\pi}=2\frac{N_c}{N}E\sin\omega t\Big]_{\alpha}^{\pi} \tag{4}$$

and substituting in equation 2 yields

$$e_{cb}=-\frac{N_c}{N_p}E\sin\omega t\Big]_0^{\alpha}-2\frac{N_c}{N_p}E\sin\omega t\Big]_{\alpha}^{\pi}+0+2\frac{N_c}{N_p}E\sin\omega t \tag{5}$$

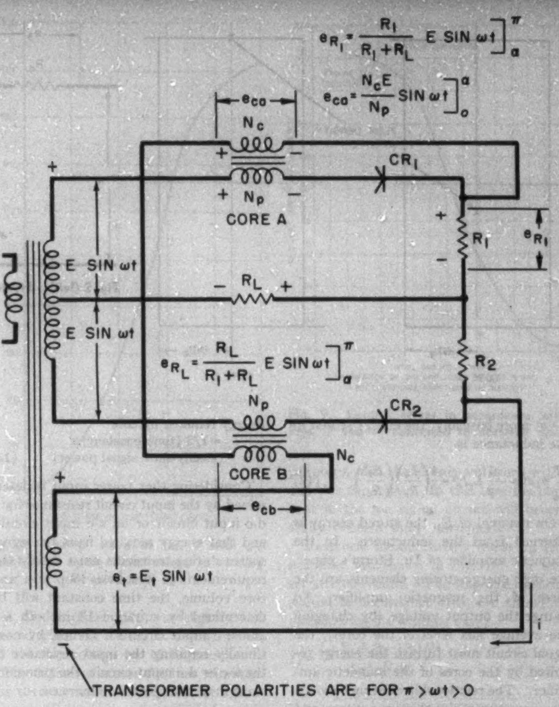
$$e_{cb}=\frac{N_c}{N_p}E\sin\omega t\Big]_0^{\alpha}+0\Big]_{\alpha}^{\pi} \tag{6}$$

Hence, core B will be resetting at the same rate that core A is setting. To meet the requirements of the **B-H** characteristic, winding C must have a current

$$i=+\frac{(NI)_E}{N_c}\Big]_0^{\alpha} \tag{7}$$

Therefore the average value of i over 1/2 cycle is equal to

$$\frac{(NI)_E}{N_c}\frac{\alpha}{\pi}$$



$e_{R_1}=\frac{R_1}{R_1+R_L}$ E SIN ωt $\Big]_{\alpha}^{\pi}$
$e_{ca}=\frac{N_cE}{N_p}$ SIN ωt $\Big]_{0}^{\alpha}$
e_{co}
N_c
+
−
CR$_1$
+
+ N_p −
CORE A
+
R$_1$
e_{R_1}
−
E SIN ωt
− R$_L$ +
E SIN ωt
$e_{R_L}=\frac{R_L}{R_1+R_L}$ E SIN ωt $\Big]_{\alpha}^{\pi}$
R$_2$
N$_p$
CR$_2$
N$_c$
CORE B
+
e_{cb}
$e_t=E_t$ SIN ωt
i
TRANSFORMER POLARITIES ARE FOR π > ωt > 0

Fig. 1. Full-wave amplifier circuit

During the half-line cycle where $\pi<\omega t<2\pi$, a similar process takes place but the current is reversed:

$$i=-\frac{(NI)_E}{N_c}\Big]_{\pi}^{\pi+\alpha} \tag{8}$$

The fundamental frequency of i is $\omega/2\pi$. This derivation shows that the rectified average value of i is related to the average value of e_{R_L} by the relationship

$$\frac{\text{average value of } e_{R_L}}{\text{rectified average value of } i}=\frac{\frac{1}{\pi}\int_{\alpha}^{\pi}\frac{R_L}{R_1+R_L}E\sin\omega t\,dt}{\frac{(NI)_E}{N_c}\frac{\alpha}{\pi}} \tag{9}$$

Taking into account the iR drop caused by this exciting current will not invalidate the analysis if an excitation frequency emf is introduced in series with e_t to furnish an emf equal to the iR drop. The value of this emf determines the output of the amplifier. This affords a means of controlling the output with an excitation frequency signal. Fig. 3 shows a circuit with an a-c signal voltage and a method of biasing added.

Time Constant

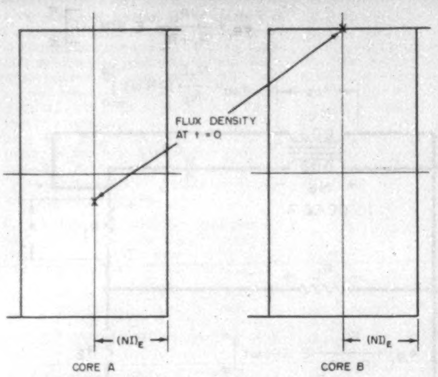
Dr. Storm[1] presents an equivalent analysis of a low-control Z, d-c input magnetic amplifier using square loop core materials. His conclusions are similar to equation 9.

In the d-c amplifier, the control circuit is effectively an R/L (reactance-inductance) circuit, as shown in Fig. 4. Assuming a step input voltage to this circuit, circuit analysis shows that the volt-time integral of the inductance is

$$\int e_c dt=-E_c\frac{L_c}{R_c} \tag{10}$$

The steady-state current through L_c is

$$I_c=\frac{E_c}{R_c} \tag{11}$$



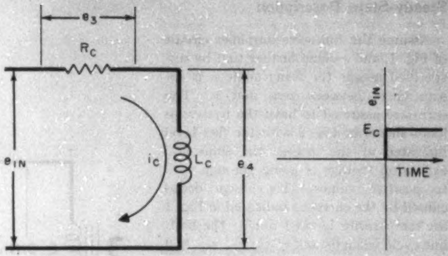


Fig. 2 (left). Hysteresis loops in units of flux density and ampere-turns

Fig. 4 (above). R/L circuit

Or, more basically, the energy stored in the inductance is

$$\int_0 p_L dt = -\frac{1}{2} L_c I_c^2 = \frac{1}{2}\left(\frac{L_c}{R_c}\right)\left(\frac{E_c^2}{R_c}\right) \qquad (12)$$

On removal of E_c, the stored energy is returned from the inductance. In the magnetic amplifier of Dr. Storm's paper, the only energy-storing elements are the cores of the magnetic amplifier. To change the output voltage (by changing the average flux level of the cores), the signal circuit must furnish the energy required by the cores of the magnetic amplifier. The relationship existing between the core energy signal source power and time constant, as stated in equation 12, may be restated as in equation 13.

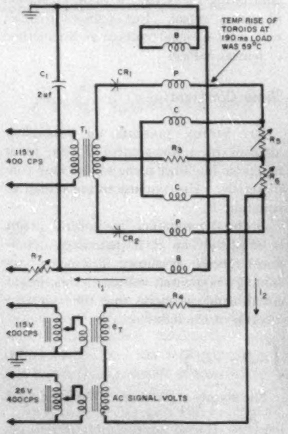


Fig. 3. Amplifier used to obtain data

Energy required by core
$$=1/2 \text{ (time constant)} \times \text{(steady-state signal power)} \qquad (13)$$

Considering that power input is determined by the input circuit resistance for a d-c input circuit or an a-c input circuit, and that energy required from the signal source during transients must satisfy the requirements of hysteresis loop area and core volume, the time constant will be determined by equation 13 in both a-c and d-c input circuits. Hence, by continually reducing the input resistance of the a-c or d-c input circuit, the time constant may be expected to increase.

Combination A-C and D-C Input Circuits

In a magnetic amplifier incorporating both a-c and d-c input circuits (neglecting control circuit iR drops), no coupling will exist in steady-state operation. Hence both types of signal circuits can be placed on the same amplifier without interaction. Then, by superposition, the firing angle (and therefore the output voltage) will be related to the signal currents by the following equation; see Fig. 3:

$$N_B \text{ (average value of } i_1 + N_C \text{ rectified average value of } i_2) = -(NI)_E \frac{\alpha}{\pi} \qquad (14)$$

where the core characteristics of Fig. 3 are described in Fig. 2.

When both a-c and d-c signal circuits are present, the time constant should be the sum of the separate time constants, just as if several d-c input circuits were used.

Measurements

An amplifier was assembled in accordance with Fig. 3. The toroid specifications were as follows:

Orthonik 0.001 tape
0.75 inside diameter by 1.125 outside diameter by 0.375 height
Winding P: 1,720 turns, no. *36* AWG (American Wire Gauge); resistance, 120 ohms
Winding B: 250 turns, no. *36* AWG
Winding C: 24 turns, no. *22* AWG

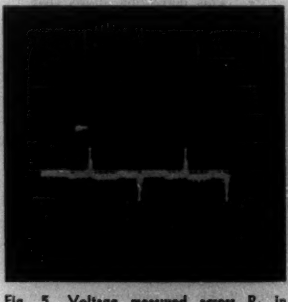


Fig. 5. Voltage measured across R$_4$ in Fig. 3 with no a-c signal voltage

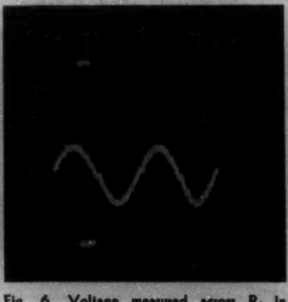


Fig. 6. Voltage measured across R$_4$ in Fig. 3 with 0.7-volt a-c signal

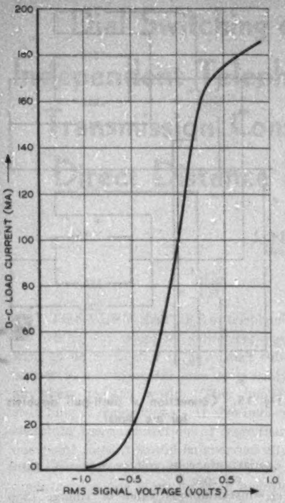


Fig. 7. Control curve of Fig. 3, current in winding B≅3.4 milliamperes

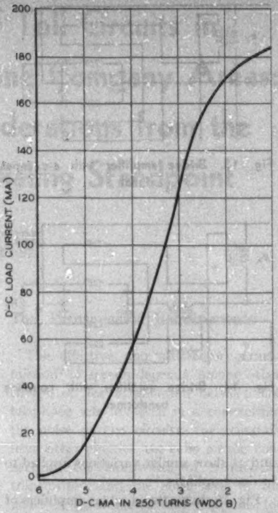


Fig. 8. Control curve of Fig. 3 with a-c signal circuit open, direct current in winding B varied with R_5

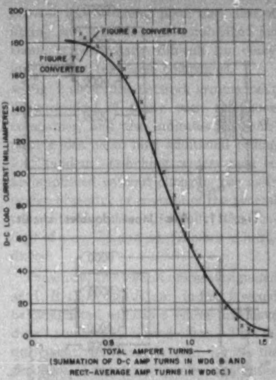


Fig. 9. Control curves in accordance with Figs. 7 and 8, converted to common units

CR_1, CR_2 = silicon diodes, 500-volt peak inverse voltage and 0.1-ampere forward current

R_5, R_6 = 20-ohm potentiometers set at approximately 15 ohms

R_3 = 400 ohms

R_4 = 1 ohm

T_1 = 400-cycle-per-second power transformer; 230-volt output center-tapped

R_7 = 50K potentiometer

C_1 = 2 μf (microfarads)

With the circuit adjusted in accordance with the steady-state description and with zero a-c signal voltage, the voltage waveform across R_4 was photographed, and is shown in Fig. 5. Increasing the a-c signal voltage to 0.7-volt rms produced the current waveform shown in Fig. 6. The gain of the oscilloscope was constant for both figures. The rms current in Fig. 6 is 20 ma (milliamperes). The a-c input circuit impedance is then 0.7 volt/20 ma = 35 ohms. This agrees roughly with a d-c resistance measurement of 31.5 ohms made on the a-c input circuit.

Figs. 7 and 8 describe the gain of the amplifier with a-c and d-c input signals respectively. One curve in Fig. 9 is a replot of Fig. 8 recording the signal in ampere turns. The other curve was found by summing the a-c and d-c signal ampere turns in accordance with equation 14. The correlation in Fig. 9 verifies equation 14. Empirically it was found that the amplifier gain was slightly improved when the a-c signal voltage phase led the line voltage by 15 degrees.

The capacitor across the bias windings was used to make the induced voltage of the resetting winding a section of a sinusoid. This is primarily for explanatory purposes, and the circuit will work without this capacitor. In this case the current in the a-c signal circuit will necessarily contain more harmonics.

Response measurements made on the amplifier are shown in Fig. 10. R_7 was set at 7,200 ohms. The response time expected in accordance with d-c theory is (considering d-c winding only)

$$T_c = \frac{I_L(R_L+R_p+R_s)N_b}{I_B R_B N_p 2f}$$

$$= \frac{0.1(535)250}{1.6 \text{ ma } (7{,}200)(1{,}720)(800)}$$

$$= 0.845 \text{ msec (millisecond)}$$

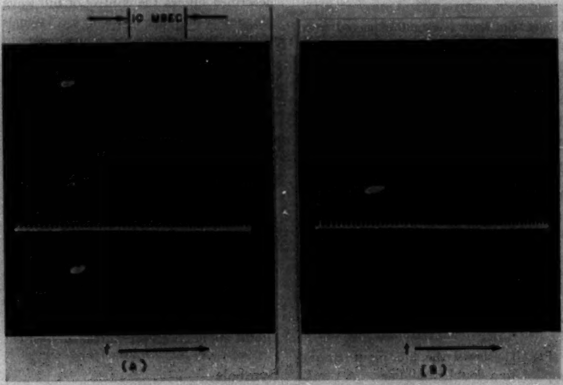


Fig. 10. Response of amplifier to an a-c step input voltage

A—Circuit in accordance with Fig. 3. Step input voltage = 0.3 volt. D-c load current change, 15 to 50 ma. Average Tc scaled from picture≅6.3 msec

B—Circuit in accordance with Fig. 3, except for the use of C_1 = 0.1 μf. Step input voltage = 0.2 volt. D-c load current change, 15 to 50 ma. Average Tc scaled from picture≅4.3 msec

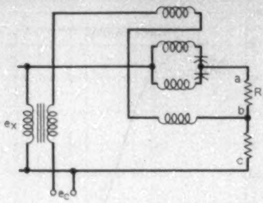


Fig. 11. A-c input doubler circuit

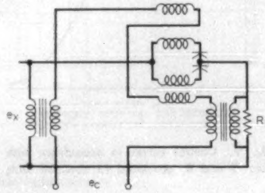


Fig. 12. A-c input doubler circuit with coupling transformer

The bias power needed to accomplish this is

$$I_B{}^2R_B=(1.6\times10^{-3})^2\,7{,}200 =18.4 \text{ milliwatts}$$

The a-c signal circuit power needed to accomplish the same change of output would be (see Fig. 8)

$$\frac{(\text{a-c signal voltage})^2}{\text{a-c signal circuit resistance}}=\frac{(0.5)^2}{33}=7.55 \text{ milliwatts}$$

Hence the time constant expected from the a-c input circuit would be

$$\frac{18.4}{7.55}(0.845 \text{ msec})=2.06 \text{ msec}$$

Therefore the combined time constant would be

$$2.06+0.845=2.905 \text{ msec}$$

The effects of the capacitor C_1 have been ignored; hence the calculated time constant is low.

The driving voltage is probably greater in Fig. 10(A) than in (B) because of nonlinearities in the control curves.

Other Variations of the Circuit

Fig. 11 shows the a-c input circuit applied to a doubler amplifier. A transformer could be used instead, as shown in Fig. 12. This variation permits reduced resistance in the signal circuit and the convenience of isolation between signal and excitation circuits. Figs. 13 and 14 show similar variations applied to bridge amplifiers.

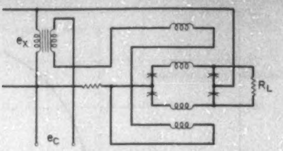


Fig. 13. Bridge amplifier with a-c input signal

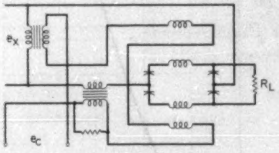


Fig. 14. Bridge amplifier with coupling transformer

Fig. 15 shows two doubler amplifiers of the type shown in Fig. 10, connected to form a push-pull amplifier. It will be noticed that the transformers used to introduce the excitation frequency voltage are bucking and are therefore superfluous. Many other circuit configurations are possible, but need not be mentioned here.

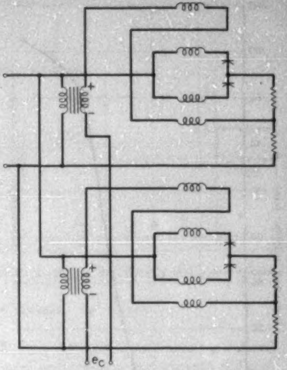


Fig. 15. Connection of push-pull doublers for a-c signal

Conclusions

The magnetic amplifier input circuit described allows the control of full-wave magnetic amplifiers with a-c signals so that the performance is analogous to d-c input amplifiers. Hence the power efficiency of full-wave amplifiers is realized. Also the power gains achievable are much higher than for intrinsically 1-cycle response amplifiers.

This input circuit permits the control of higher-powered magnetic amplifiers with low-power, class *A*, silicon transistors. This permits the advantages of each to be combined without the use of a demodulator. Summing of a-c and d-c signals is also possible.

It is applicable to many varieties of single-sided or push-pull amplifiers, as shown by the several examples of typical circuit applications presented.

Symbols

α = firing angle in radians
ω = line frequency in radians per second
t = instantaneous time
e_{ca} = instantaneous voltage across N_c turns on core *A*
e_{cb} = instantaneous voltage across N_c turns on core *B*
e_{R1} = instantaneous value of voltage drop across R_1
e_{R2} = instantaneous value of voltage drop across R_2
e_t = instantaneous voltage induced in transformer winding
N_c, N_p, N_B = number of turns on windings *C*, *P*, and *B* respectively
E = peak value of line voltage
R_1 through R_7, R_c, R_K = ohms resistance
i = instantaneous current; see Fig. 1
E_t = peak voltage induced in transformer winding indicated in Fig. 1
$(NI)_s$ = coercive force in ampere turns; see Fig. 2
L_c = inductance in henrys
i_c = instantaneous current; see Fig. 4
I_c = steady-state direct current; see Fig. 4
E_c = direct voltage; see Fig. 4
P_L = instantaneous product of voltage across an inductor and current through it
i_1, i_2 = instantaneous currents; see Fig. 4
CR_1, CR_2 = rectifiers
T_1 = transformer
C_1 = capacitor
P, B, C = toroid windings
R_p = power-winding resistance
E_B, I_B = corresponding changes in d-c bias voltage and bias current
e_x = excitation voltage
e_c = signal source voltage

References

1. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE LOOP CORE MATERIALS, H. F. Storm. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 629–40.

2. RESPONSE TIME OF MAGNETIC AMPLIFIERS, L. J. Johnson, S. E. Rauch. Hufford Machine Works, Inc., Los Angeles, Calif.

3. THE INFLUENCE OF MAGNETIC AMPLIFIER CIRCUITRY UPON THE OPERATING HYSTERESIS LOOPS, Harold W. Lord. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 721–28.

4. RESPONSE TIME OF MAGNETIC AMPLIFIERS, E. L. Harder, W. F. Horton. *Ibid.*, vol. 69, pt. II, 1950, pp. 1120–41.

Dial Switching of Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint

J. N. PETRIE
MEMBER AIEE

BY THE VERY NATURE of telephone traffic, any telephone subscriber may wish to be connected to any other subscriber in a continental area or possibly anywhere in the world. This implies switching of connections. It also implies switching these connections at more than one point since it would be economically impracticable to concentrate all switching operations at one point for even a single metropolitan area. With subscriber line switching on a decentralized basis, it naturally follows that switching centers themselves must be switched in order to accomplish economical interconnection of any two telephone subscribers.

Requirements for switching have been set up on the basis of a systematic plan which has been in effect for many years and is most commonly known in the commercial telephone communications industry as the "general toll-switching plan."[1] This paper will not go into the details of this plan except to point out that the present version of it presupposes assigning every telephone exchange into one of five classes. Fig. 1 shows the class designations and relationships as well as the different paths which might be involved when subscribers in two different offices communicate with each other. If these are class *5*, a total of eight switches or nine circuits (not counting the subscriber lines) may be involved. The operation of telephone connections involving several switches presents some challenging problems. This paper will be confined to only one of these problems, that of meeting the transmission requirements.

The Transmission Requirements

The effective and satisfactory transmission[2] of speech depends, among other things, on the attenuation between the telephone sets involved in a connection, the noise on the circuits, the crosstalk from other circuits, the echo on the connection, the bandwidth of frequencies transmitted, and the noise level at the point of transmission and reception (or more precisely, the acoustic conditions surrounding the speaker and the listener). In modern long-distance telephone circuits, the major factors contributing to the quality of a connection from a transmission point of view are attenuation and echo. These factors are to a large extent dependent on the "return loss" obtainable at 2-wire switching points. Return loss is primarily a measure of the degree of balance existing at the hybrid coils (which are invariably involved in these connections), and is therefore an indication of the magnitude of the echo which may be expected. On connections having low attenuation the echo tends to be even more noticeable. The problem of complying with transmission requirements of the general toll-switching plan, and also of fitting into classifications established for them, is particularly difficult for many independent telephone companies.

Fig. 2 shows the network arrangement quite typical of many such telephone companies, and also indicates this company's idea as to how additional coding for "tandem" switching centers may be accomplished without changing the basic design objective or concept of the general toll-switching plan. Basically, the concept of a no. *5A* and no. *5B* office is introduced.

Presently, the several offices designated as *A*, *B*, *C*, *D*, and *E*, etc., in Fig. 2 would be designated as class 5 offices and have direct trunks to a class *4* office. This requirement, however, is contrary to the practices of many companies who desire to operate what might be termed "county-wide" networks and, for economic reasons, would not wish to provide direct circuits from the several offices such as *A*, *B*, *C*, *D*, and *E* to the class *4* office. Alternatively, all offices classed above *A*, *B*, *C*, *D*, etc., (end offices) would necessarily have to be upgraded from a class *5* to a class *4*, a class *4* to a class *3*, etc., in accordance with the current general toll-switching plan. This would be a distinct hardship, especially on the Bell System.

To switch circuits at a tandem office, designated as the no. *5-A* office, presents a number of important and difficult transmission problems when working into the nation-wide DDD (direct distance dialing) network. Also some engineers believe that service rendered through a PBX (private branch exchange) is fundamentally similar to that rendered through "tandem" switching offices. Certainly PBX's constitute a special transmission problem unless they are to have long distance trunks direct to toll centers

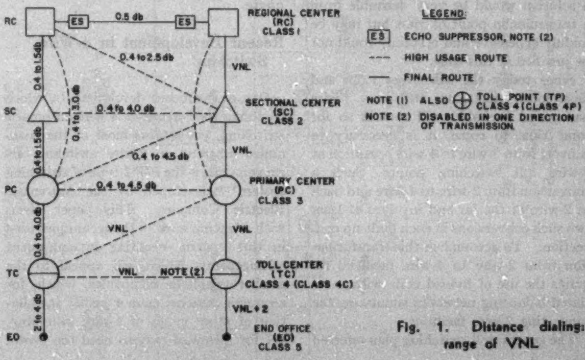


Fig. 1. Distance dialing: range of VNL

Paper 58-386, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted October 4, 1957; made available for printing April 30, 1958.

J. N. PETRIE is with the Automatic Electric Company, Northlake, Ill.

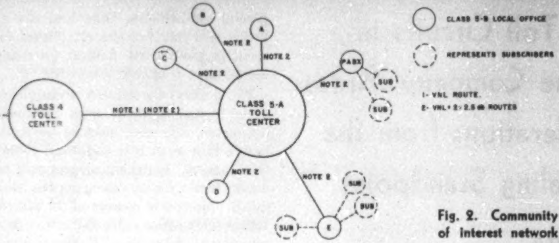


Fig. 2. Community of interest network

rather than to their own end offices. From a transmission standpoint, PBX does constitute the same problem as the outlying offices previously designated by the letters, *A*, *B*, *C*, and *D*, and is therefore in the same category.

The Nature of Telephone Circuits

Telephone conversations are bidirectional or duplex in nature, i.e., conversation must be possible in either direction over a single connection and reversible at will without any delay or noticeable interference to the talkers. To meet this requirement, it would be highly desirable to have available separate 1-way circuits for each direction of transmission, each equipped with a transmitter at one end and a receiver at the other, thus providing for 2-way conversation at all times, and at a loudness level limited only by the amount of amplification introduced in the respective 1-way transmission paths. Carrier channels, which are used extensively for long-distant telephone purposes, do provide bidirectional full duplex (or what are more commonly known as 4-wire) circuits. By the provision of suitable 4-wire local trunks and subscriber loops the objective could be achieved of an over-all 4-wire transmission path. Such a solution would be very desirable from a transmission point of view but may be unduly expensive and certainly could not be justified at this time.

Since today the subscriber loop, and most local trunks, are still on a 2-wire basis, and probably will remain so for some time to come, it is necessary to convert from 2-wire to 4-wire operation at 2-wire toll switching points. Such a conversion from 2-wire to 4-wire and back to 2-wire at the far end involves at least two such conversions in each built-up connection. To accomplish this transformation from 2-wire to 4-wire facilities requires the use of hybrid coils with associated balancing networks simulating the connecting 2-wire facilities.

The general toll-switching plan referred to in the foregoing is based on extensive use of 4-wire long-distance circuits, and 4-wire switching whenever economically feasible. No doubt, ultimately, all of the long-distance circuits and trunk circuits will be carrier derived and 4-wire switching will become more popular, so that the desirable characteristics of 4-wire operation can be achieved. Actually, 2-wire switching offices will be in use for years to come; however, it would still be desirable to take advantage of the favorable characteristics of 4-wire toll-switching while retaining the 2-wire local circuitry and equipment.

Requirements for 2-Wire Switching

Experience indicates that when a 27-db (decibel) return loss between circuits is attained at a 2-wire switching office, for all intents and purposes the performance will be comparable to that of a 4-wire switching arrangement. However, this involves a considerable initial expenditure and continuing effort and expense to maintain the degree of precision and control over the plant facilities found necessary to retain the 27-db return loss objective. At points where the 4-wire facilities are actually switched to 2-wire end links, a 15-db return loss will usually be adequate.

Recent Development in 4-Wire Switching

One of the recent developments, which combines the features of 4-wire automatic switching, yet retains most of the economies inherent in 2-wire switching arrangements, is the *FW-1* type of switching system[3–6] developed by the Automatic Electric Company. This paper deals with adapting some of the techniques used in this system, resulting in equipment arrangements which will appeal to the smaller telephone companies, which, for economic reasons, cannot justify installation of other types of 4-wire switching, yet for technical reasons need the 4-wire performance inherent in this switching arrangement. Such a solution involves using 4-wire toll-switching techniques not dependent on return losses at any class of office and not requiring a manual toll switchboard. To appreciate the significance of this arrangement it is necessary to review several possible conditions.

Fig. 3(A) shows diagrammatically what the problem is in terms of present typical plant facilities, attenuation, and return loss considerations. The return loss at the no. *4* office will be of the order only of 15 db. It should be noted that calls from a no. *4* office to the no. *5B* office (condition *A*) do not meet DDD minimum transmission requirements of 4 db as presently established, since no gain can be switched in at no. *5A* office.

Fig. (3B) shows a somewhat similar arrangement ultilizing 4-wire switching at a no. *4* office which results in an improved return loss figure; however, again no provision is available for gain switching at the no. *5A* office; therefore the transmission loss from no. *4* to no. *5B* end office (illustrated by condition *A*) exceeds the 4-db objective. Condition *B* shows that better than the minimum transmission requirements are attained when the call terminates at the no. *5A* office.

Fig. 4(A) indicates what might be expected if 4-wire switching were provided at tandem switching centers (designated as the no. *5A* office) with carrier-derived channels connecting it to the no. *4* office. In this instance, 2-wire switching is assumed to be used at the no. *4* office, resulting in a return loss of 15 db. It is particularly significant to note in condition *A* that high return loss and, consequently, gain switching is possible at the no. *5A* office. This will permit VNL (via net loss) switching between the no. *4* and no. *5A* offices and effectively make a no. *4* office out of the no. *5A* office, permitting better than minimum standard transmission to no. *5B* end offices. Condition *B* shows the transmission arrangement on traffic-terminating in the no. *5A* office. The 10-db return loss indicated is predicated on the improvement obtained by insertion of the 2.5-db pads in both the line and network branches of the hybrid coil. This improvement is essentially a "by-product" derived from the "padding out" of the carrier channel from the VNL condition to the TNL (terminal net loss) condition. No change in loop gain on the carrier channel itself is involved. It should be noted that the transmission losses from the no. *4* to the no. *5A* or no. *5B* offices are substantially identical and within the transmission objectives for an end office.

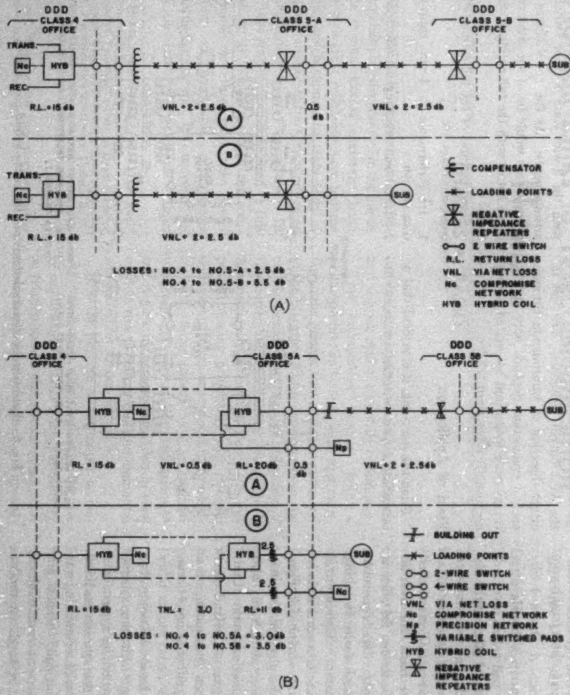

DDD CLASS 4 OFFICE
DDD CLASS 5-A OFFICE
DDD CLASS 5-B OFFICE
TRANS.
REC.
Nc
HYB
SUB
R.L. = 15 db
VNL + 2 = 2.5 db
0.5 db
VNL + 2 = 2.5 db
A
B
R.L. = 15 db
VNL + 2 = 2.5 db
COMPENSATOR
LOADING POINTS
NEGATIVE IMPEDANCE REPEATERS
2 WIRE SWITCH
R.L. RETURN LOSS
VNL VIA NET LOSS
Nc COMPROMISE NETWORK
HYB HYBRID COIL
LOSSES: NO. 4 to NO. 5-A = 2.5 db
NO. 4 to NO. 5-B = 5.5 db
(A)
DDD CLASS 4 OFFICE
DDD CLASS 5A OFFICE
DDD CLASS 5B OFFICE
Np
RL = 15 db
VNL = 0.5 db
RL = 20 db
0.5 db
VNL + 2 = 2.5 db
2.5
RL = 15 db
TNL = 3.0
RL = 11 db
BUILDING OUT
LOADING POINTS
2-WIRE SWITCH
4-WIRE SWITCH
VNL VIA NET LOSS
Nc COMPROMISE NETWORK
Np PRECISION NETWORK
VARIABLE SWITCHED PADS
HYB HYBRID COIL
NEGATIVE IMPEDANCE REPEATERS
LOSSES: NO. 4 to NO. 5A = 3.0 db
NO. 4 to NO. 5B = 3.5 db
(B)

Fig. 3. Present tandem office problems

A—Two-wire toll center switching
B—Four-wire toll center switching

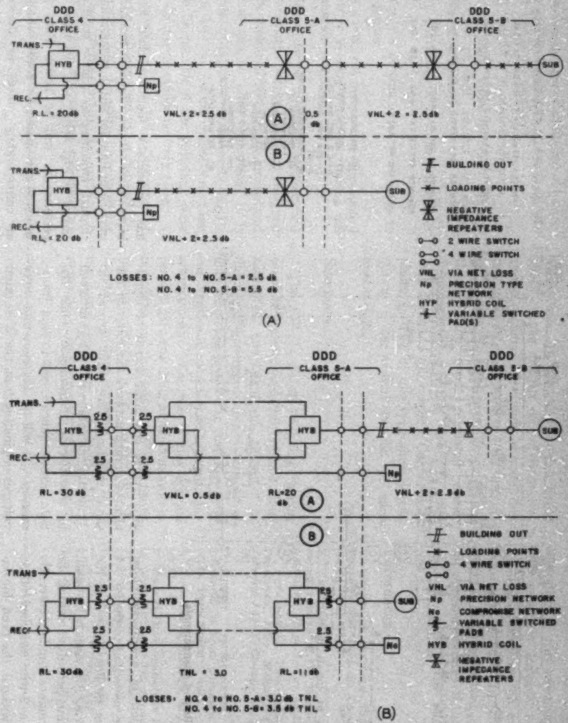

DDD CLASS 4 OFFICE
DDD CLASS 5-A OFFICE
DDD CLASS 5-B OFFICE
TRANS.
REC.
HYB
Np
SUB
R.L. = 20 db
VNL + 2 = 2.5 db
0.5 db
VNL + 2 = 2.5 db
A
B
R.L. = 20 db
VNL + 2 = 2.5 db
BUILDING OUT
LOADING POINTS
NEGATIVE IMPEDANCE REPEATERS
2 WIRE SWITCH
4 WIRE SWITCH
VNL VIA NET LOSS
Np PRECISION TYPE NETWORK
HYP HYBRID COIL
VARIABLE SWITCHED PAD(S)
LOSSES: NO. 4 to NO. 5-A = 2.5 db
NO. 4 to NO. 5-B = 5.5 db
(A)
2.5
RL = 30 db
VNL = 0.5 db
RL = 20 db
VNL + 2 = 2.5 db
RL = 30 db
TNL = 3.0
RL = 11 db
Nc
BUILDING OUT
LOADING POINTS
4 WIRE SWITCH
VNL VIA NET LOSS
Np PRECISION NETWORK
Nc COMPROMISE NETWORK
VARIABLE SWITCHED PADS
HYB HYBRID COIL
NEGATIVE IMPEDANCE REPEATERS
LOSSES: NO. 4 to NO. 5-A = 3.0 db TNL
NO. 4 to NO. 5-B = 3.5 db TNL
(B)

Fig. 4. Variable loss at 2/4-wire switching offices

A—Two-wire toll center switching
B—Four-wire toll center switching

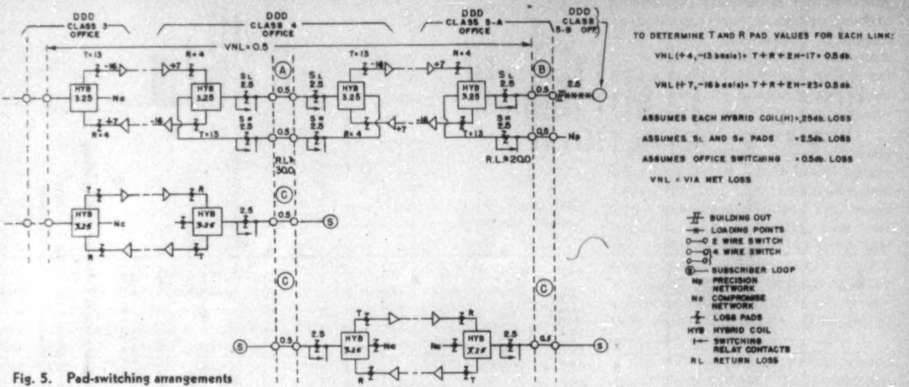

DDD CLASS 3 OFFICE
DDD CLASS 4 OFFICE
DDD CLASS 5-A OFFICE
DDD CLASS 5-B OFF.
VNL=0.5
T=13
R=4
HYB 3.25
Nc
Np
SL 2.5
SR 2.5
0.5
2.5
-16
+7
A
B
C
S
RL≥300
R.L.≥200
TO DETERMINE T AND R PAD VALUES FOR EACH LINK:
VNL (+4, −13 basis) = T+R+2H−17 = 0.5 db.
VNL (+7, −16 basis) = T+R+2H−23 = 0.5 db.
ASSUMES EACH HYBRID COIL(H) = .25 db. LOSS
ASSUMES SL AND SR PADS = 2.5 db. LOSS
ASSUMES OFFICE SWITCHING = 0.5 db. LOSS
VNL = VIA NET LOSS
BUILDING OUT
LOADING POINTS
2 WIRE SWITCH
4 WIRE SWITCH
SUBSCRIBER LOOP
Np PRECISION NETWORK
Nc COMPROMISE NETWORK
LOSS PADS
HYB HYBRID COIL
SWITCHING RELAY CONTACTS
RL RETURN LOSS

Fig. 5. Pad-switching arrangements

The application of 4-wire switching to a class *4* office, as well as to the no. *5A* offices, as shown in Fig. 4(B), has the effect of improving the return loss substantially, and thus permits exceeding the minimum performance standard currently established for a 2-wire switched office, namely 27 db, without the need for expensive office balancing. The conditions at the no. *5A* office remain unchanged from those assumed in Fig. 4(A).

Fig. 5 indicates schematically how the net loss is controlled and network switching accomplished in the *FW-1* switching equipment. Condition *A* shows 4-wire switching of two carrier channels at a class *4* office and the pad values involved. The link from the class *3* office to the no. *4* office, as well as the link from no. *4* to the no. *5A* office, would be operated at VNL = 0.5 db. However, when switched together on a 4-wire basis, the over-all loss from the no. *3* to the no. *5A* office is also only VNL = 0.5 db. This is desirable and permissible because 4-wire switching is essentially transparent. Condition *B* shows 4-wire switching of a carrier channel and a physical trunk at a no. *5A* office. It is assumed that the physical circuit is set up for operation at 2.5 db and equipped with a precision network. Condition *C* shows the pad values and compromise network when terminating on a local subscriber loop, either at no. *4* or no. *5A* office.

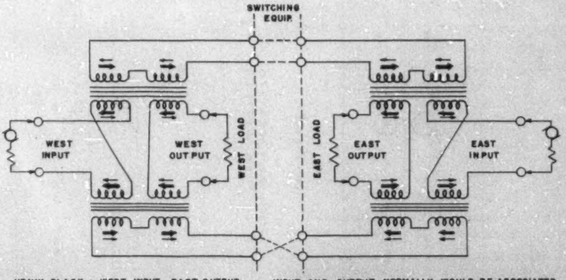
Fig. 6 shows the basic philosophy behind the "tail-eating"[7] switching scheme used in the *FW-1* switching system and used as a basis for the solution outlined in this paper. Four-wire terminating sets of the "two-repeating-coil" version are shown as the transformer action involved is more easily understood. Actually, any hybrid coil arrangement or equivalent 4-branch network functions in the same manner. The heavy arrows indicate the instantaneous voltage involved in west-to-east transmission while the light arrows indicate east-to-west transmission. It will be apparent that voice voltages originating from the West input terminals cancel out in the West output and the East input terminals, while adding in phase only at the East output. Conversely, voice voltages originating from the East input terminals appear only at the West output terminals. The loss in either direction of transmission through such a connection depends only on the transformer losses and associated wiring and usually is less than 1/2 db.

Field of Application

From the foregoing, it is evident that a tail-eating 4-wire switching arrangement has considerable merit at class *5A* or higher class offices. In each instance it may be utilized for either terminating or through switching to achieve results as shown in Fig. 4(B). On this basis, the performance will be better than that which is anticipated for 2-wire switching at corresponding offices as presently visualized for the general toll-switching plan. When used at the no. *5A* offices, the switched trunk return loss will be better than 20 db, whereas only a 15-db return loss is expected to be accomplished with the 2-wire class *4* office switching arrangements. At the no. *4* offices the return losses would be of the order of 30 db or better with the tail-eating arrangement.

Some features of the tail-eating hybrid coil connection are

1. Its performance is 4-wire equivalent transmission.
2. Initial and recurring expense (due to


SWITCHING EQUIP.
WEST INPUT
WEST OUTPUT
WEST LOAD
EAST LOAD
EAST OUTPUT
EAST INPUT

HEAVY BLACK = WEST INPUT – EAST OUTPUT
LIGHT BLACK = EAST INPUT – WEST OUTPUT

INPUT AND OUTPUT NORMALLY WOULD BE ASSOCIATED WITH A FOUR-WIRE TELEPHONE CIRCUIT (SUCH AS A CARRIER CHANNEL).

Fig. 6. Instantaneous current relationships: tail-eating hybrid connections

subsequent changes or additions) of balancing an office to achieve or retain a high return loss is not necessary.

3. It is flexible in open application.

4. It permits interconnecting terminal repeatered physical circuits or carrier channels of all makes.

5. It interconnects a variety of impedances.

6. There is automatic insertion of up to 6 db of gain.

7. It is relatively inexpensive when applied in new or existing offices not having toll transmission switches.

When network terminals are switched to precision networks which are designed for the matching of a specific facility and without specific office balancing, it is possible to achieve a distinctly higher return loss on 2-wire connections than would be possible if the switching were done on a compromise network basis.

Utilization of the tail-eating type of switching arrangement at tandem switching centers (no. *5A* office) may have far reaching implications in the application of Strowger automatic toll-ticketing equipment. It may also have a major effect on the number of toll-connecting trunks required. It should be noted that the successful application of 4-wire switching at a no. *5A* office presupposes utilization of toll grade regulated carrier channels interconnecting the no. *5A* office with the no. *4* toll center. Carrier telephone systems of this type already constitute a substantial portion of such applications.

Summary

The *FW-1* type of switching scheme provides features and transmission characteristics equivalent to those of a directional 4-wire switching system but at only a nominal increase in cost over present commonly used 2-wire toll-switching arrangements. The use of this scheme permits gain switching at tandem switching offices, and thus solves a difficult transmission problem for community-of-interest telephone systems when it is integrated with the nationwide DDD network. The conversion of most existing offices to this switching arrangement will generally involve replacement of existing intertoll switch trains and their associated wiring. In those situations where a new installation is contemplated the superior features of this arrangement certainly justify the slight additional cost.

References

1. NOTES ON DISTANCE DIALING. American Telephone and Telegraph Company, New York, N. Y., 1956.

2. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, H. R. Huntley. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 670–76.

3. FUNDAMENTAL PRINCIPLES OF INTERTOLL SWITCHING, Imre Molnar. *Automatic Electric Technical Journal*, Chicago, Ill., Dec. 1956, pp. 50–53.

4. RECENT DEVELOPMENTS IN FOUR-WIRE SWITCHING OF LONG DISTANCE TELEPHONE CIRCUITS, Imre Molnar. *Ibid.*

5. TOLL TRUNK CIRCUITS, E. J. Glenner. *Ibid.*, pp. 71–75.

6. RECENT DEVELOPMENTS IN 4-WIRE SWITCHING OF LONG-DISTANCE TELEPHONE CIRCUITS, Imre Molnar. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 722–33.

7. CORD CIRCUIT REPEATER FOR USE BETWEEN FOUR-WIRE AND TWO-WIRE CIRCUITS, George Crisson, A. F. Rose. *U. S. Patent 1,465,298*, Aug. 21, 1923.

Factors Affecting the Use of Over-the-Horizon Links in Telecommunication Networks

CHARLES A. PARRY
NONMEMBER AIEE

THE SCATTER propagation mechanism has placed a new tool in the hands of the communication engineer. Long spans with great reliability have become feasible, and a supporting technology of high-powered transmitters, huge antennas and multiple receivers is available. In order that scatter circuits be used to full advantage as integral parts of the communication structure, more than determination of acceptable performance standards is required. It is necessary to assess the extent to which these links can be incorporated in existing facilities, and can supplement existing techniques.

This examination requires both a philosophical and technical approach. Up to the present time little has appeared in the literature which considers both the capabilities and the optimum utilization of the scatter link as a communication tool. A considerable amount of detail has appeared dealing principally with specific behavior patterns or component development.[1,2] Some papers have generalized with respect to performance limitations,[3,4] but in general these need revision in view of more recent information and cannot be considered for serious design.

It is the purpose of this paper to discuss the factors which influence the use of multichannel scatter links, particularly in relation to national telecommunication network planning. Scatter is a specialized technique and each individual possibility has to be examined on its own merit. It should be borne in mind, that for this reason, this paper is offered only as a guide toward final evaluation.

Generalization

It is necessary at this point to discuss the use of generalizations as they apply in the present discussion. In order to evaluate the scatter circuit in terms that will permit its utility to be assessed relative to other communications facilities, it is necessary that certain generalizations be made. This technique is, of course, quite common in the art. However, scatter presents a particular problem on which there should be a note of caution. In those cases when it is possible to make use of the generalizations, it is also likely that other communication methods would be suitable. By its very nature, therefore, scatter lends itself to those areas which present special problems. It is of a particular nature rather than general. This implies that while the fundamentals of the scatter method must be known, each circuit requires treatment case by case, which negates against the use of generalizations. It should be understood therefore, that although it is necessary to use generalizations in order to discuss the subject intelligently, basic recommenda-

---

Paper 58-383, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 4, 1958; made available for printing March 17, 1958.

CHARLES A. PARRY is with Page Communications Engineers, Inc., Washington, D. C.

tions laid on such a framework may prove to be quite misleading. This point should be retained when evaluating the ensuing text. It is because of this difficulty in generalization that philosophical concepts related to implementation must be evolved, the better to establish the specific areas for which scatter may be suited. Furthermore this may indicate modus operandi which will open a greater market to the use of this technique.

Evaluation of Preferred Techniques

When a decision must be made as to the implementation of a given service—and the world is used here as a generic to cover a link, circuit, technique, or whatever may be involved—certain judgment factors must be established for guidance, otherwise the results must be completely arbitrary. Admittedly the arbitrary judgment of an experienced engineer, may yield what ultimately proves to be correct results. The difficulty with this process is that no behavior pattern is established, and subsequent decisions cannot be made with equal certainty of success.

These judgments are an attempt to define the acceptable limits for the important criteria of the service, so that the final decision ascribes proper "weighting" to all relevant facts. It should be realized, however, that many of the details which must be considered reduce to subjective evaluation.

For instance, if a given administration considers that it is too difficult to train staff to operate a given service, who can prove this incorrect? And how can the importance be assessed other than by the ultimate opinion of individuals. If it can be shown that the service increases revenue so that a school can be established for training, how can this new fact be weighted other than by human judgment? There are many facets of a communication network which must be considered in the same way. The net results are the best attempts to correlate quantitative data with these subjective factors.

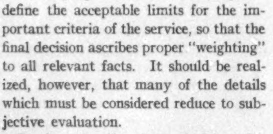


Fig. 1. Units required in subscriber-to-subscriber link. Indicates the relatively small part played by radio link in large network

The principal parameters on which judgment should be made in the evaluation of a communication service are as follows:

1. Compatibility
2. Performance
3. Price of design error
4. Strategic implication
5. Operation and maintenance
6. Economic assessment
7. Administration

Applicable factors are discussed separately in this text, together with certain detailed aspects of the decision problem which appear to be related particularly to the scatter circuit.

Compatibility

It is always difficult for the engineer who has been deeply immersed in a development of great promise, to find there is a lack of immediate universal acceptance, and that the industry is not "revolutionized." For an understanding of this it is necessary to realize the way in which an art develops.

When a new technical field is opened, there is usually a spectacular rate of progress as new ideas become practical and are adopted. As the art develops, so too does the net invested capital. Trends which would obsolete this capital must offer something in return and mere "better performance" rarely has significance. Thus it becomes more and more difficult for an established industry to be revolutionized. Conservatism is proportional to invested capital and is the balance wheel for the overindulgent claims of new ideas. An established field automatically imposes restrictive limits on new developments if they are to be integrated with the existing services; and no development has any value unless it can be utilized. Thus there is a golden rule which states, "The utility of any service can only be measured in relation to what already exists." In order to be acceptable there must be a minimum conformity with these restrictions. This is referred to under the general heading of "compatibility." The radio industry has recently had this problem brought sharply into focus by the requirements of "compatible" color television.

It is important to emphasize, however, as was apparent in this latter instance, that intrinsic cost and performance capability alone are not sufficient criteria for the determination of compatibility. Also there are sometimes factors involved which lie deep in human psychology. It is the net assessment of both human and quantitative factors which ultimately determine the extent to which a new service is acceptable.

In general it can be stated that compatibility requires a study of a number of loosely related factors. These are as follows:

1. Economics
    a. Basic cost of implementation
    b. Cost of obsoleting other services
    c. Maintenance and other service costs
    d. Revenue
2. Integration with existing facilities
    a. Performance limits
    b. Additional equipment needed to interconnect to existing circuits
    c. Increased maintenance problems
    d. Logistics
    e. Extension of present services
3. Social technology
    a. Ability physically to implement new service
    b. Ability to stand capital investment with possible long-term amortization
    c. Staff training
    d. Schooling for public acceptance and use
    e. Administration of new service

The scatter circuit is only a small part of the complete communication link. An attempt to indicate this is shown by the typical subscriber to subscriber circuit of Fig. 1. This has the advantage that new problems have proportionately less impact on the total task; the disadvantage is that specialized conditions may be demanded by a disproportionate part of the entire system. Furthermore, scatter is only one of several available operational techniques.

From these facts it may be deduced that the conditions requisite of compatibility for the scatter link will be quite extensive, embracing many fields of endeavor. A detailed study patterned on the basis of the factors outlined in the foregoing list is usually essential.

Performance

It is not the purpose of this paper to analyze the details associated with the evaluation of performance characteristics. Nevertheless it is essential to outline these broadly in order to assess the capability of the scatter technique relative to other methods. Furthermore, new commercial opportunities may become apparent in this process.

For the following discussions, median path loss had been obtained from the latest data made available by Norton,

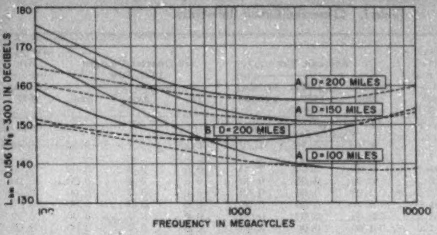


Fig. 2. Relationship between frequency and transmission loss

D = distance over smooth spherical earth
N_s = refractive index in N units
A = curves based on 28-foot parabolas
B = curves based on 60-foot parabolas

Center of antennas 30 feet above earth. Dotted curves obtained by letting Norton's frequency gain function equal zero. Margin added for 99% long-term fading. No diversity and no margin for short-term fading

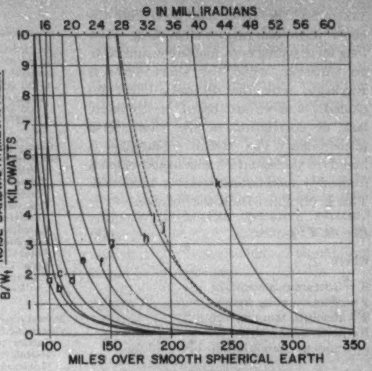


Fig. 3. Performance index B/W_t versus distance for various conditions of reliability, based on equation 1. Center of antenna in all cases is 30 feet above earth. Frequency 2,000 mc/sec. Noise figure 8 db. Total line loss 5 db. The curves are as follows:

Curve	Antenna Diameter, Feet	Diversity	Reliability, %
a	10	dual	95–99
b	10	quadruple	99–99
c	18	dual	99–99
d	10	dual	95–90
e	18	dual	95–99
f	28	dual	99–99
g	18	dual	95–90
h	28	quadruple	99–99
i	28	dual	95–90
j	60	dual	99–99
k	60	quadruple	99–99

Data apply to single-link system

based on his original paper.[5] Diversity improvement is based on ratio squared combiners as outlined by Altman and Sichak.[6] Long term fading is established by the most recent information evolved by Norton in an unpublished report. Medium aperture coupling loss is determined by Stara's method.[7] A total line loss of 2.5 db (decibels) at each end of the system has been allowed in all cases.

FREQUENCY OF OPERATION

Up to the present time, the variety of frequencies used has arisen, not so much from a study of operation parameters, as the necessity to make use of equipment already in production. As more knowledge becomes available this situation is likely to alter.

Early data seemed to indicate a relatively simple relationship between transmission loss and frequency.[3] More sophisticated analyses have shown that a complex relation exists. Fig. 2 has been drawn on the basis of present information to show this dependence, and certain conclusions can be drawn from this data. In so far as transmission loss can be used as a guide, the "general" optimum frequency centers around 2,000 mc/sec (megacycles per second). There is little difference to be seen over the range 1–4,000 mc/sec for a wide range of conditions. For short paths, higher frequencies tend to become more useful, while for long paths, particularly those with high sites, low frequencies may be preferable.

However, the choice of frequency will be influenced by factors other than transmission loss. Not the least of these is the ease with which allocations in the required range can be obtained from controlling authorities such as Federal Communications Commission, or International Telecommunications Union. The lower frequencies, below say 1,000 mc/sec are likely to suffer in this respect.

At lower frequencies, the antenna structure required to give gain equal to that of large parabolas, can be much simpler and there can be a substantial reduction of cost. Since the lower frequencies tend to be favored by long spans, this question of cost may offset the problem of frequency allocation.

Perhaps even more significant is the pattern of interference which exists around the scatter installation. This has been under study by responsible organizations for some time. The growing use of scatter renders the problem more and more acute. Yeh[8] has established a method for calculating the interference field based on some of the methods evolved from these study groups. The high power usually involved means also that harmonics and spurious signals, even though well down, may be radiated with high absolute power. Front to back antenna ratio assumes considerable significance: This factor tends to increase with frequency in practical installations and may become of considerable importance at low frequencies.

Tentative recommendations appear to be taking the trend that scatter and microwave links on the same system are mutually exclusive. Certainly, before scatter is used, the interference pattern for all radiated frequencies should be determined. This is one powerful reason which should favor higher frequency allocations; the total interference field can be substantially reduced. Certainly it would appear preferable to keep the scatter frequencies above the microwave frequencies (on a common system) if this is at all possible.

Summing up, the choice of frequency is influenced by:

1. Transmission loss
2. Interference field
3. Possibility of allocation of frequency
4. Relative cost
5. Availability of equipment

CIRCUIT RELIABILITY

The term circuit reliability[9] is used here to refer to the effects of propagation. In order to indicate a specified outage effect, the fading ranges associated with both long term and short term fading have been indicated separately. For instance a curve 90%–99%, means that the long term fading margin has been taken for the 90% abscissas, and the short term margin for the 99%. The long term fading is of course considered to be gaussian, while the short term is Rayleigh, with the diversity effect included. Curves are based on complete lack of correlation of the short term phenomena. The circuit reliability is based on the signal P_t which just reaches frequency modulation receiver threshold. This is obtained from the relation

$$P_t = FKTB\epsilon \tag{1}$$

where

K = Boetzman's constant
F = effective noise figure
T = absolute temperature
B = noise bandwidth
ϵ = frequency-modulation threshold factor

We can put $P_t = W_t(A)$ where W_t is the transmitted power, and A the total transmission loss factor. This includes medium aperture loss and fading margins. It then is convenient to express the performance in terms of B/W_t; this has been shown in Fig. 3. The data here are based on 2,000 mc, and a noise figure of 8 db. A suitable figure for ϵ can be obtained from Stumper[10] and Crosby;[11] both authors indicate 10 db is adequate although values different from this are used by some engineers. Ten db has been used in Fig. 3.

It is convenient to refer to B/W_t as the "performance index" of the system since it clearly relates to the essential definitive parameters of the link. If we accept as a guide the concept that $d\ (B/W_t)/d_\theta$ should be a small, it would appear from Fig. 3 that B/W_t should preferably be <1. However, it should be noted that for low values of power, quite important distances and reliability can be achieved and this must be kept in mind for commercial development of the scatter industry. It is also seen that extremely important reduction in the ordinate can be achieved by a reduction in the reliability demanded. So far no literature has given a guide toward acceptable values for outage time; even the CCIR (International Consultative Committee on Radio)[12] has so far avoided the problem.

The following limits are suggested on a tentative basis:

	Long Term	Short Term
High grade telegraphic circuits	99%	99%
Medium grade telegraphic circuits	99%	90%
Telegraphic circuits with error-correcting techniques	95%–90%	99%–90%

Table I. Observed Circuit Reliability

Source of Circuit Outage	Average, Hrs Sep–Dec 1956	Average, Hrs Jan–Mar 1957	Average, Hrs Apr–Jun 1957	3rd Quarter Outages 1957, Hrs Jul	3rd Quarter Outages 1957, Hrs Aug	3rd Quarter Outages 1957, Hrs Sep	Average, Hrs Jul–Sep 1957
West-to-East Circuit							
Primary power system*	3.40	4.30	1.12	0.45	6.8	0.55	2.60
Electronic equipment:							
Carrier terminal equipment	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Modulation equipment	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Monitron, rf transfer	0.02	0.00	0.00	0.0	0.0	0.00	0.00
Transmitting equipment	0.02	0.00	0.00	0.0	0.0	0.00	0.00
Receiving equipment	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Antenna systems	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Propagation	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Total Circuit Outage	3.44	4.30	1.12	0.45	6.8	0.55	2.60
Total Time Period	732.00	720.00	728.00	744.0	744.0	720.0	736.00
Long-Term Circuit Reliability, %	99.53	99.40	99.85	99.94	99.09	99.92	99.65
East-to-West Circuit							
Primary power system*	3.40	4.3	1.12	0.45	6.8	0.55	2.60
Electronic equipment:							
Carrier terminal equipment	0.00	0.00	0.00	0.35	0.0	0.00	0.12
Modulation equipment	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Monitron, rf transfer	0.11	0.05	0.00	0.0	0.0	0.00	0.00
Transmitting equipment	0.06	0.00	0.25	0.0	0.0	0.00	0.00
Receiving equipment	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Antenna systems	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Propagation	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Total circuit outage	3.57	4.35	1.37	0.8	6.8	0.55	2.72
Total time period	732.00	720.00	728.00	744.0	744.0	720.0	736.00
Long-term circuit reliability, %	99.51	99.40	99.81	99.89	99.09	99.92	99.63

* Outages represent total for both terminals for each time period.

Separate study indicates that voice circuits are not so vulnerable to the outage effect as telegraphic circuits, although further investigation is certainly required.

For voice channels the following limits are suggested:

	Long Term	Short Term
High grade circuits (This corresponds to an outage of less than one hour in a year)	99%	99%
Commercial circuits	95%	99%
Limiting or military circuits	95%	90%

It should be recognized however in connection with the last limit, that military requirements often dictate that telegraph circuits be connected over the voice channel: in the absence of adequate error-correcting equipment the limits shown would be too low.

It is perhaps useful to note that the reliability which is reasonably feasible, is still in excess of values ordinarily associated with high-frequency circuits. This leads to the serious consideration that the lower grade scatter circuit may serve a useful commercial purpose in bridging the gap between the long span high-frequency technique and the short-span very-high-frequency and microwave technique. Certainly there are many areas where excessive limits of reliability are uncalled for, yet better reliability is required than is possible with high-frequency circuits, and longer spans than is possible with very-high-frequency circuits.

EQUIPMENT RELIABILITY

Equipment reliability in practical installations is quite high. There are two basic reasons for this.

1. There are fewer actual operating units, involved in the scatter link than the conventional microwave system. There does not appear to be significant loss in reliability as the power of a given type of unit is increased, so the over-all reliability of the scatter link improves.
2. Maintenance is generally more effective. The large θ installation is always attended, while it is almost mandatory to have unattended microwave stations. In the former case, when breakdowns occur they can be corrected quickly since the station is staffed. Problems associated with dust infiltrating and poor weather conditions are greatly minimized, since much potential trouble is eliminated due to the normal maintenance procedure of keeping the building in good order.

Power failures, which contribute significantly to total unreliability figures can usually be quickly overcome. As discussed later, the scatter system, particularly when part of a national network, is likely to have its stations near to metropolitan areas. This usually insures that a relatively reliable local supply is available.

The high order of reliability for the typical installation is shown by Table I. The propagation outages in this particular chart tend to be a little misleading, since

Table II

No. of Channels	f_m, Kc/Sec.	Loading Factor L_f, Db	Factor A	White Noise Correction S/N/S/D	p_L/p_m with S/N = 40 Db	B, Mc/Sec	Peak Deviation d, Mc/Sec. for S/N = 40 Db	B/d	Reduction in Threshold for Modulation Index = 2 and d = B/2, Db
120	552	15	22.4	17.6	3.2	3.6	2.4	1.48	2.1
60	300	13.5	21.3	16.1	3.4	2.04	1.41	1.45	2.3
48	252	13	21	15.6	3.6	1.63	1.2	1.36	1.8
24	108	11.5	18.8	14.1	4.2	0.9	0.61	1.47	3.1
12	60	10.5	17.3	12.6	4.4	0.530	0.41	1.31	3.4
	S/N (30 db)				2.0	0.240	0.184	1.31	0
5	24	9	14.8	10.7	5.6	0.270	0.185	1.46	4.5
3	16	8	14	9.5	5.8	0.185	0.135	1.37	4.6
		3*	19	16	4	0.128	0.09	1.42	3
Average								1.41	

* Value for typical commercial circuits.

the circuit was designed with more than adequate margin in this respect. The data are presented primarily to show that the circuit reliability compares very favorably with other communications systems.

SIGNAL-TO-NOISE RATIO

The absolute signal-to-noise ratio will depend upon a number of factors, and the minimum requirements have not yet been generally agreed upon. Certain aspects of this problem have appeared in the literature and some suggestions have been made.[9] There is reason to believe that satisfactory performance for national toll circuits can be achieved with a signal/weighted noise ratio of 40 db exceeded for 99% of the busy hour. For non-integrated circuits, the operational value of 34 db for 95% of the busy hour has been suggested. Lowest practical commercial design would be about 30 db for 95% of the busiest hour and for reasonable military circuits, probably 15 db for 99% of the busy hour. The activity factor during the busy hour should be determined by the class of service rather than that obtained in large exchanges operating at maximum traffic capacity.

It can be readily shown that the signal to weighted thermal noise ratio on the top channel of a frequency-modulation system at threshold, is given by[9]

$$S/N : \frac{BK^{-1}}{2bL_f}\,\epsilon\delta(M^2) \tag{2}$$

where

B = noise bandwidth of receiver
b = channel spacing
ϵ = crest factor at threshold (taken as 10 db)
K = weighting factor (taken as 0·5 for 4 kc channels)
δ = improvement due to the use of pre-emphasis and de-emphasis (taken as 2 for the top channel)
M = modulation index
L_f = load capacity factor

This may be put in the form

$$S/N = \left(\frac{s}{p_m}\right)^2 \left(\frac{p_L}{p_m}\right) \underbrace{\left(\frac{f_m}{2b}\right)\frac{2K^{-1}\epsilon\delta}{L_f}}_{\text{factor } A} \tag{3}$$

The first two bracketed terms correspond to similar factors used by Medhurst[13] in his evaluation of the signal to nonlinear noise S/D associated with bandwidth reduction.

There is no simple solution to equation 3. However, a satisfactory solution can be made by assigning a definite value to S/D. This is then plotted in terms of (p/p_m) and $(s/p_m)^2(p_L/p_m)$. If S/N is specified, (p/p_m) can be found and the associated band-width and deviation determined.

Both f_m and L_f depend on the number of channels, so that these computations must be performed for each particular system. Values of f_m based on the use of CCIF (International Consultative Committee on Telephony)[14] frequency allocations for the higher capacities are given in Table II.

Medhurst's value of S/D is based on the use of white noise test signal. A correction must be made to this in order to obtain the expected signal/nonlinear noise ratio S/N_1, in the operative system.[9] This correction is associated with the statistics of the multichannel signal, and it can be shown that

$$\frac{(S/N_1)}{S/D)} = \frac{bn^{1-\alpha}}{X(1.5)(10^9)} \tag{4}$$

where n = number of channels and X and α are constants determined by the slope of the load capacity curve. A recent study (a paper in the course of preparation by the author) has been made of this, and values of L_f suitable for toll quality circuits are also given in Table II. For these values $x \simeq \alpha \simeq 0{\cdot}5$ and the correction obtained from equation 4 has also been added to the table. This is based on a traffic activity factor of 0.25.

With S/N known, an appropriate value of S/N_1, can be determined. With n known, S/N can be established and equation 3 then solved.

In order to arrive at working numbers it was considered that the nonlinear noise from this effect should be 10 db below thermal noise. A further allowance of 8

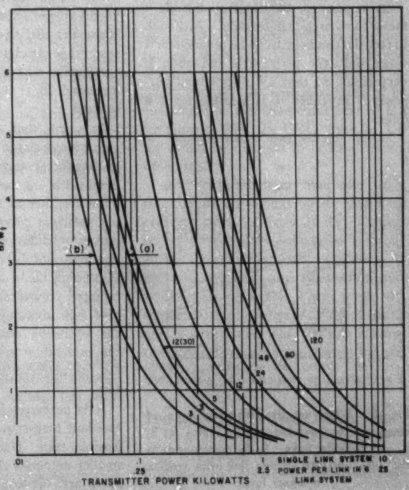


Fig. 4. Performance index versus transmitter power, with number of channels as parameter. All curves except a and b are based on 40-db signal to weighted thermal noise at threshold, with an activity factor $\tau = 0.25$. Bandwidth used includes 10% safety margin. Data based on B in Table II. These curves can be used in conjunction with Fig. 3 to establish reliability and θ values

A—30 db s/n
B—Low activity as in commercial system

db was made for the 6 links of a reference system. Hence for a signal-to-thermal-noise ratio of 40 db, $S/N_1 = 58$. S/D for the 120-channel system is then about 40 db.

Although some exception might be taken to these approximations, the bandwidth effect is such that even relatively large changes in S/N_1 will not call for any substantial change in B.

The end product of all the arithmetic associated with the determination of B is given in Table II. On the basis of this approach it can be expected that the nonlinear noise due to bandwidth restriction will be more than 10 db below threshold thermal noise for 99% of the busiest hour. This applies for at least 1–6 links.

The average value of B/d is 1.41, which suggests that the recommended ratio might well be 1.5 for all practical circuits.

Fig. 3 has been drawn with an ordinate B/W_t. Using equation 3 as outlined in the foregoing, B is determined for a given number of channels, so that the required power in terms of n and B/W_t, is easily established. This is shown in Fig. 4. Figs. 3 and 4 may now be used to obtain the transmitter power for a given reliability, distance, and number of channels for a 40 db signal-to-thermal-noise ratio at threshold. All the curves of Fig. 4 have been drawn to include a 10% increase in the bandwidth B indicated by Table II, as a safety margin.

The effect of reduced requirements is shown for 3 and 12 channels and is seen to reflect substantially on the required power.

As mentioned, these curves include a factor for 6 circuits in tandem. Although in fact the bandwidth can be reduced for a single link, the permissible reduction is very small, so that for all practical purposes B in Table II is suitable for 1–6 links.

Many commercial links require a relatively small number of channels. In the case of the scatter installation this requires a disproportionately high cost for the radio equipment. This is one of the factors which is adversely affecting commercial application. Part of the difficulty is associated with the fact that although channel capacity must be provided, the usage rate is relatively low. Concentrators have been described[15] which permit a much better matching of the capacity to the actual load although they do not appear to be yet available for radio circuits. In principle these use channels to provide searching functions, and then connect a multiplicity of lines to unused trunks. Thus if the expected activity is 0.1 the density could be increased by 2.5 times before reaching an activity of 0.25. This is the maximum value found in large exchanges[16] and therefore is not likely to be exceeded under conditions of full circuit utilization.

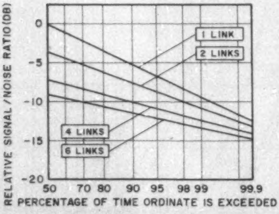


Fig. 5. Effect on signal-to-noise ratio of cumulative uncorrelated gaussian fading for 2, 4, and 6 similar links. 0 db corresponds to the ratio obtained in one link with median rf signal level

The application of this technique might well repay examination in the scatter field. Consider for instance a low density 24-channel system. A given reliability (constant B/W_t) could be achieved by a reduction of some 3 or more times the transmitter power, depending on actual values chosen. Alternatively, if the power is assumed to be fixed, a smaller antenna installation is possible. A figure of 5 channels has been chosen as representing a reasonable compromise between required performance and possible utilization of concentrators, in low capacity applications.

The same technique can be also applied to larger capacity systems of course. However, the need for larger capacity systems is associated with a greater social complexity and the utilization factor is increased so that concentrators cannot be used.

It is true that a reduction in activity factor is associated with a reduction in the loading factor, and therefore influences the bandwidth, as seen from equation 3. However, due to the problem of obtaining sufficient phase linearity in the frequency-modulation system, the bandwidth cannot be reduced beyond a certain point, i.e., a minimum bandwidth is required regardless of activity factor, so that full advantage cannot be taken of the operative condition. Consequently the preferable technique is to maintain the full activity by virtue of concentrators.

It is perhaps of interest to note that excellent long span circuits appear to be feasible with much lower power and smaller antennas than has been common in the art. It would appear that there is a definite place for the 100-watt installation, and this does not appear to be the minimum figure permissible.

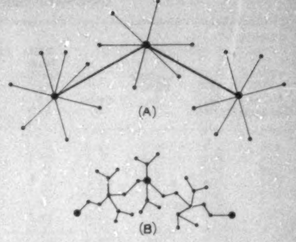


Fig. 6. Radial network required with scatter and the spine network possible with microwave, when both systems must serve intervening terrain

A—Scatter
B—Microwave

COMPANDORS

The prime advantage of the compandor of course is to raise signal-to-noise ratio and thereby permit what is basically, a lower quality circuit. An advantage of as much as 20 db can be obtained by this process.[17] However the compandor does not affect the circuit reliability, which is primarily a function of P_t/B. There is some reason to believe that slightly less outage time is permissible with the use of compandors, because loss of signal produces effects due to compandor action which tend to reduce intelligibility. However the system can be designed on a minimum bandwidth basis, i.e., $B = 4f_m$, since nonlinear noise-to-signal ratio is also improved.

Comparing this with B in Table I and with reference to equation 1, it is seen there is a reduction in the threshold. The extent of this reduction can therefore be used in conjunction with Fig. 3, to estimate the transmitter power for a given B/W_t. This technique makes the 5-channel system with line concentrators look particularly attractive. A 100-watt transmitter gives $B/W_t = 3$. The reduction due to minimal bandwidth is 4.5 db, making B/W_t now equal 1.06. An 18-foot dish with dual diversity and 95-99% reliability then gives an effective range of over 140 miles. This is a significant coverage with components which are within a reasonable price range. The cost of the radio equipment is now comparable with multiplex and such a system would appear to be economically attractive. Such a system will carry quite good standard voice-frequency telegraph circuits, since the circuit reliability is reasonable and the the channel S/N without compandors is adequate for this purpose.

The combined techniques of compandors and concentrators appears to be an approach worth a great deal more investigation by the industry.

FICTITIOUS CIRCUIT OF REFERENCE

Although the CCIR has not yet defined a fictitious circuit of reference for scatter systems there seems no fundamental reason why the concept could not be retained. Although it is very unlikely that a circuit of the reference length could ever exist with scatter, there is the real requirement that it may be part of such a circuit. From Fig. 3 it would appear that a reasonable distance for a link could be 250 miles. On this basis it might be suggested that the scatter reference circuit should consist of 6 homogeneous sections similar to the CCIR[18] suggestion for certain multichannel radio relays.

Since a typical microwave link is about 30 miles, the single scatter link in the reference circuit would be equivalent to eight microwave links.

The effect of fading over these links must be considered. Long term fading over a distance of 250 miles, may be considered as having a gaussian distribution with standard deviation of 4 db. Although some correlation may be expected over the several links, particularly from that component of the field associated with seasonal changes, it seems reasonable at this stage to establish parameters on the basis of uncorrelated fading. Curves for the combined distribution due to the several links based on Mack's[19] work are given in Fig. 5. Since the design figures used for the high quality circuits previously was based on 99% long term reliability, it would appear that the relative signal-to-noise ratio for a single section of the reference circuit should be about 4 db higher than for the overall system. Since higher reliability per section is also required, it can be expected that the power indicated in Fig. 3 should be increased by 4 db to make the data suitable for the single link of a fictitious reference system. Alternately this increase can be also obtained with the use of larger antennas than those indicated for the specific curves of Fig. 2.

ANTENNA LINES

A few words should be added concerning antenna lines. For the large θ installation, these lines may be quite long. With wideband systems the possibility of nonlinear noise due to echo can become quite serious since the noise power is proportional to the fourth power of line length.[20] A good match at the antenna can reduce this, but with high power links this may not be an insignificant factor.

Applicability

The scatter link has the basic advantage of being effective over relatively large spans without losing the principal advantages that have been normally associated with radio relays. In order to investigate the possibilities which scatter offers, it is necessary firstly to consider it on a functional basis. This will determine its appropriateness to specific communication requirements. The conditions which tend to favor scatter may be divided broadly into three categories.

1. Strategic: where the avoidance of the intervening stations represents a major step toward security.
2. Environmental: where the terrain prohibits the use of conventional radio relays, for instance,
   a. Desert
   b. Ocean
   c. Very rugged mountains
3. Commercial: where the economies permit scatter to be proved in
   a. Links providing special service for a particular organization, viz. mining, oil, and utility companies
   b. Links which are part of national or quasi-national networks

It is not useful to discuss the strategic application of scatter links. Strategic importance is rarely associated with economic factors, and may dictate developments which do not extend into the commercial field. This implies that growth of the scatter industry is likely to be overdependent on the state of world tension, budget allocations, and current military policies. Consideration along these lines is therefore not a part of this paper. It is also difficult to obtain a relative evaluation of scatter in those cases where a commercial radio relay is virtually excluded. Of necessity the economics of such a link have to be judged per se. Much the same remarks apply when the link provides a special commercial service.

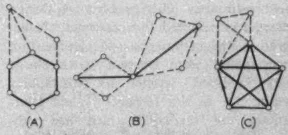


Fig. 7. Some methods by means of which alternative routing is possible for overflow traffic. Dotted lines are for high loss probability circuits

A—Ring system
B—Star system
C—General regional interconnection method

As part of a national or quasi-national network, however, the scatter link lends itself to particular functional arrangements. The purpose of a network is to service all significant areas of the community. Since scatter avoids contact with intervening terrain, areas between scatter stations must be served by a radial system. This is quite different from the spine structure possible with microwave, shown in Fig. 6. This has the advantage that the single scatter link carries the entire traffic between regions and it is possible that sufficient revenue can be obtained to amortize the investment.

Alternative routings for the overflow of traffic is an essential part of any network structure. Typical, though by no means exhaustive, methods are indicated in Fig. 7. It is difficult to see where the scatter link can fit into such a pattern. The long span implies either an equally long nonscatter overflow route, or that the scatter itself is for the overflow. In the first case it would appear that other linkage methods over the same route are feasible, and if economically suitable for overflow traffic, it is difficult indeed to see how a prime scatter link would be justified. In the second case it is doubtful if overflow revenue could justify its presence. This would seem to relegate scatter in the national network to acting as part of a backbone, without alternate routing facilities. This in turn would demand high reliability. If standby or power change over switching were to be considered as a fundamental part of the equipment, then the CCIT limit of 150 ms would be dictated in order to avoid drop outs.

Also it would seem that only main regional center should be so linked, in order to permit sufficiently high channel capacity to obtain requisite revenue. Interconnection between satellites, or between satellites and regional centers would seem unlikely. These would not require a multiplicity of trunk circuits, and it is unlikely that the low channel capacity demanded would justify the cost of the scatter link.

The Price of Effective Design Error

It is basic to the concept of system planning that an estimate should be made of expected performance. There is always, of course, a difference between this prediction and the results achieved in the field. This is referred to here, generically, as the "effective design error" although it is apparent that the differential is not, in the precise meaning of the word, due to an error. This effective error accumulates for the following reasons:

1. Malicious intent
2. Incompetence
3. True errors in the data used by design engineers
4. Statistical deviations in empirical material
5. Approximations
    a. Mathematical
    b. Graphical
6. Variations in equipment performance
7. Estimates where there is inadequate information

Modern business could not function if items 1 and 2 were at all significant and the discussion therefore encompasses only the remainder. A substantial effort is made in general to reduce the importance of item 3 so that ordinarily only items 4–7 are of prime importance.

The degree to which an attempt is made to reduce the error depends to a great deal on the risk involved, risk usually being measured in terms of reputation and capital cost.

However, even where the risk is relatively small, there is, concomitant with the scientific method, an attempt to reduce this to reasonable proportions. In fact one of the principal present commercial difficulties is to persuade engineers to provide designs which do not call for performance in excess of risk error.

The minimum value to which the effective error can be reduced, at the present state of the art, is not well established. Even with the best care possible it does not seem reasonable to expect that path attenuation, including fade margin, can be predicted closer than ±3 db even when supported by field data. The figure is, in fact, somewhat less than the most probable error indicated by Norton's[4] work. To this must be added the performance variations associated with equipments. Usually engineers must design to one of two broad conditions. Either equipment is purchased on the open market, or else it is manufactured in the organization of which he is a part. In the first case, it is usually impossible to determine statistical performance deviations. In the second, if need arises, these cannot only be determined, but where necessary, can often be reduced by new engineering developments.

For the first condition mentioned it does not seem reasonable to reduce the net effect of all of the variations expected from the various components of a radio link to better than ±3 db. This figure probably can be reduced in the second case, but the extent of this reduction is not readily apparent and is not likely to be very substantial.

On this basis it would seem that the reasonable total minimum effective error is ±6 db. This implies that under the most careful conditions, it is not reasonable to expect the performance in situ to be better than ±6 db from the calculated design and this figure might well be regarded as a minimum design margin. When there is considerable "scatter" in the information, then a value in excess of this can be expected.

The effective error becomes of increasing significance as the basic cost of an installation rises, and it is the true error component of this—which does not become apparent until after initial tests are made—that causes the most difficulty. It is necessary to reduce this to zero, then changes must be made to the installation for this purpose. In the scatter link, the cost rises steeply as the scatter angle increases, due to the use of high power and large antennas. An installation which already has 10-kw transmitters, 60-foot dishes, and quadruple diversity, imposes a prohibitive cost if only a 3-db true error must be corrected.

This question of risk is usually more serious in the case of scatter circuits than for equivalent microwave circuits. This is because the price per db is normally much greater for the former. This implies that, as the basic cost increases, so does the need for a smaller design margin.

Most organizations require a performance guarantee of some kind. In order to minimize the risk an additional design margin is incorporated into the over-all system plan. However, as the basic cost of the link rises, smaller margins are desirable in order to reduce the net capital cost estimate. This problem of design margin is reflected to a considerable degree in the commercial development of any system, and because of the large risk involved, scatter is particularly vulnerable in this respect. When guarantees are involved which are substantially inviolate, there is a further margin added for "unexpected" deviations in performance. Consequently, the tighter the guarantee, the higher the estimated basic cost. It is, perhaps, not realized by those who insist on performance standards, the extent to which not only themselves as a customer, but the industry as a whole is hindered by this problem. There is need for an understanding of this difficulty and a more realistic assessment of real performance requirements demanded for any service. For instance, the tendency to use Holbrook and Dixon[16] curves for purely commercial circuits is open to question. Data relevant to this problem have appeared[9] and more widespread consideration of the design parameters involved is required.

Strategic Considerations

The factors which influence the strategic importance of a circuit are rarely measured in conventional economic terms, and must be judged on a per se basis. It cannot be expected that engineers who have had no direct contact with military problems should have a clear understanding of the strategic impact of any communications method; guidance must come from those skilled in these techniques. The scatter link appears to have some important advantages from the strategic viewpoint.

1. The long span permits a circuit with a minimum number of links, so that there are fewer stations to protect.
2. Contact can be made over very difficult terrain.
3. The required size of the typical installation means protective measures do not impose disproportionate cost.
4. The long span possible permits contact to be established at points which might otherwise be considered impossible.

The high cost of scatter links with large scatter angles is likely to limit to a large extent such circuits to purely military applications. In order to obtain the maximum distance consistent with cost, performance requirements should be tailored to absolute need. A significant parameter is the loading required. There is some evidence that even for full capacity working, the limits used are too high. Signal-to-noise ratio suitable for military working is much lower than required for toll operation and 15 db weighted is probably an acceptable figure. The emphasis is on circuit reliability, particularly since so many circuits are telegraphic; outage time can interfere seriously in this case, while it is not so important in the case of voice. These aspects imply that the bandwidth can be substantially reduced over that normally required, the distortion associated with this being no longer a deterrent factor. On the basis of a 3-channel system using the minimal bandwidth mentioned, and 99%–99% reliability, it would appear from Figs. 3 and 4 that the limiting value

Vol.
Pt. 1
195

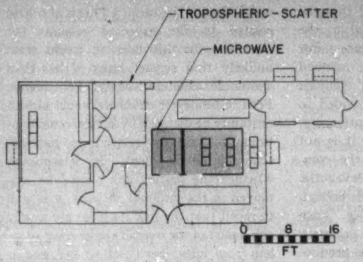


Fig. 8 (left). Relative sizes of microwave and scatter (10 kw) stations

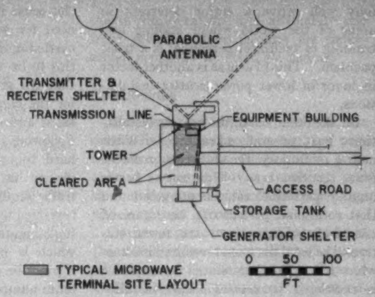


Fig. 9 (right). Relative sizes of microwave and scatter (10 kw and 60-foot dishes) sites

of scatter angle is well over 80 milliradians.

The use of minimal performance standards for strategic circuits generally precludes their use as part of specific toll networks. In a similar way, when part of the national network is justified by carrying strategic circuits, the higher performance requirements limit the range of the military circuit. It is here that a conflict arises which can only be reconciled by the administrations involved.

Significance of the Terrain

Every communication technique must be evaluated with respect to the terrain its spans. This is not simply a matter of the propagation phenomena, but is related to what are primarily factors concerned with logistics and installation. There are of course, cases when the terrain is quite impossible for conventional radio relays but is readily accessible by scatter. The question then is not so much one of relative evaluation as a judgment of its value per se. On the other hand, there are cases where a microwave link can be installed with relative ease, and scatter cannot be proved in. Between these two limiting situations exists the need for a specific assessment of the relative advantage to be gained from the use of scatter. It is probable that a more thorough study of this will eventually yield a critical value of θ versus distance, indicating the "break" point between scatter and microwave links. Certainly it would seem that as θ increases for a given distance, it might be expected that the intervening terrain would be progressively worsening, weighing the balance in favor of scatter.

The scatter link can offer specific advantages which are usually associated with two factors: 1. The long span avoids intermediate stations, and 2. it is frequently possible to have the ends of the link near metropolitan areas. This is particularly true when the scatter circuit is part of a national network. The very nature of the basic requirements of the network virtually dictate that this will be so.

Centers which can be connected by scatter can usually achieve this with a single span. This implies that the terminal installation can be placed relatively near a populated area. This usually means that construction facilities are much more readily obtainable.

The physical area required for the high power installation is quite large compared to the typical microwave station. This is depicted in Fig. 8. However, there is usually considerable freedom in choosing the site, since θ is relatively insensitive to quite large movements of either terminal. The buildings required are much more elaborate for scatter, but this is offset, of course, by the number of stations. The relative problem is indicated by Fig. 9 which shows a typical microwave terminal building, relative to a large θ scatter installation.

Most of the factors which must be taken into account in determining the influence of the terrain can be debated in a similar fashion. The principal factors which must be weighted in the relative evaluation of scatter and microwave systems connecting the same end points are outlined in the following.

1. Survey. Relatively little work over the intervening terrain required by scatter.

2. Antenna direction. Direction of antennas must be precise since relocation can be very costly. However, total number of sitings is less.

3. Location. (Scatter terminals probably near metropolis.) Easier to obtain constructing agencies. Basing and outfitting work crews simpler.

4. Number of stations. Fewer with scatter. Total road length less. Many sites may be difficult of access in microwave system. Logistics during installation phase simplified for scatter.

5. Buildings. Scatter stations more complex, but fewer are required. Total floor area probably greater than microwave, but smaller quantity may simplify construction problems.

6. Site. Scatter terminal requires much larger area than microwave. Usually greater freedom in site choice due to θ insensitivity.

7. Refurbishing stations. Probably easier for scatter particularly when located near metropolis; in severe weather the situation may be under better control.

8. Path loss tests. Scatter equipment likely to be bulky. Tests require relatively long period on single path, so that in a given time the sample of data obtained is more meaningful than the same time spent over several different microwave paths. Also as the number of paths increase, the probability that one may be in error increases.

Operation and Maintenance

The operation and maintenance of a complex unit cannot be divorced from the associated factors of logistics and training. The maintenance of the scatter unit is not simply a matter of supplying trained men. They require special knowledge. In high power installations a whole new field must be covered, dealing with such things as pressured lines, high power diesel generators, and other items peculiar to the installation. Furthermore, this problem is not simply a question of increasing the extent of knowledge. The cost of poor maintenance and operation increases rapidly with the capital cost of the installation. The risk involved to personnel also rises very quickly. These things are not minimized by learning more. Individuals with an increased sense of responsibility are required, and it is well known that the proportion of personnel available decreases rapidly as the demanded level of responsibility rises. Consequently, both knowledge and caliber beyond that required for the conventional radio relay are required. This single consideration may well prove a major deterrent in areas of the world where staffing and training of technical personnel is a basic problem. This of course is another factor in favor of lower power scatter installations.

On the other hand, the scatter installation has several advantages. When near a metropolis, the staff can be drawn from the city area. This means better quality personnel can be attracted and that rotational periods can be organized effectively. Since there are fewer stations, the total personnel required is somewhat less. Further, since maintenance is restricted to central locations, difficulties which are concomitant with bad weather servicing are minimized. Similarly problems associated with dust infiltration and the effects of rain, snow and ice are greatly reduced.

Also related to this question is one of logistics. When the scatter link is associated with other radio links, additional spares have to be kept for the one link alone. This may not seem a great difficulty, but does pose an additional problem of storing, cataloguing, and replacing the special items necessary. This would be greatly simplified if the scatter frequencies could be similar to those of the associated radio links, since it is conceivable that much of the equipment would be common. However, if the question of interference precludes this, as is usually the case, there may be quite large differences in basic equipment and special supplies are necessary. As against this, refurbishing the commensurate microwave system, often occurs at almost inaccessible sites, and in bad weather this may outweigh the disadvantages which are attributable to the scatter installation.

Economic Evaluation

The need for disassociated groups of people to communicate with each other is bound up in an intricate way with their ethnic and economic interdependence. When the intervening terrain is such that scatter seems preferable to other accepted methods of communication linkage there is the implication that this terrain is inimical to the social intercourse of the groups. Thus it must be realized that the scatter link will join semiautonomous groups which have become sufficiently sophisticated to reduce their isolation. Since the scatter link primarily eliminates contact with the intervening areas, it must be inferred that the need has arisen for the connected groups to enter into some form of economic relationship. In this sense it may be considered that the need for scatter is proportional to the need for social interdependence. For this to be valid it must also be assumed that the groups are sufficiently complex for the communication link provided to assist and develop this interrelationship. Following this line of reasoning it is not hard to consider that unless the complexity has reached a certain level, the traffic required could not provide enough revenue to justify its presence, since fundamentally it is an economic necessity which is involved. From this premise it must be assumed that there is a minimum number of channels required of the scatter link before it can be proved in. There is no clear pattern as to what this might be, but it is certain that a scatter circuit cannot be justified simply because it would be cheaper than a commensurate microwave link. It is also apparent that this "break through" number will depend both on the degree of social complexity and on the absolute need involved.

It should also be apparent that the degree of group isolation must be related in some way to the value of θ; since large θ over a short distance implies rugged terrain, while large θ over smooth terrain implies great distance. Both of these factors tend to isolate the groups, and therefore their need for interrelationship is even more dependent on social complexity. We may expect therefore that the number of channels needed by the groups might tend to diminish with increasing θ. On this basis, it would seem as though there would be critical θ, beyond which it would be expected that the number of channels demanded would not support the cost of providing them. It is also apparent there must be a minimum θ which is such that intermediate stations would be desirable, mitigating against scatter. Hence it would appear that there is a range of θ values over which the economic factors associated with human development are contributory toward the use of scatter links.

These discussions will be recognized as particularly applicable to the use of the scatter link as part of a national or international network.

In establishing a break-through number it should be considered that the present day structure of any telephone system is based on the cost with which the required components may be constructed. To be acceptable, any single link of the system cannot disturb this economic balance. In general, it would appear that for a radio link to be justified the cost of the multiplex equipment must outweigh that of the ratio. This is of course related to the expected revenue per channel. On this basis it would seem unlikely that scatter links of less than about 24 channels would be feasible. Fewer number of channels would almost centainly be dictated by special considerations not simply related to national economic development. In determining whether the scatter link should be employed in the national network, it is well to recall that the link itself is only a very small part of the overall engineering problem from subscriber to subscriber, and as part of a complex network, comprises only a part of the total cost. This is indicated to some extent in Fig. 1 which shows the separate parts of the subscriber chain.

For these reasons, the engineering difficulties and basic costs themselves are of relatively minor consequence. Judgment has first to be made on the functional value which has been discussed under Applicability. The scatter link may easily be less costly than a commensurate microwave link over the same path, but it may be necessary to make contact with areas along the route and this cannot be done if the scatter link is employed.

Mention has been made of the fact that the justification for scatter is also related to absolute need of separate groups to communicate. This may be related to two factors not yet discussed. Firstly there is one of self preservation, that is, the circuit is required for its strategic value. Secondly there is the need in which the economic assessment is determined not by the interconnected groups themselves but by outside agencies which find financial return from contact with the groups. There are many such cases of course. For instance mining and oil companies may establish links to areas from which wealth can be obtained, but which, due to their own lack of development have not yet made use of this wealth. Summing up these considerations then, it would seem that the prime justification for the use of scatter is deeply rooted in the question of socio-economic interdependence. The categories of circuits for which scatter is feasible are as follow:

1. Circuits to be integrated into the national or international plan. This is dependent on the need for the groups to establish economic bonds.
2. Circuits other than 1.
    a. Needed for their strategic value or for purposes of safety
        1. Military
        2. Weather
        3. Aviation
        4. Emergency in disaster

b. Special circuits established primarily by groups outside those which are actually connected.
   1. International
   2. Commercial organizations such as mining, oil and power companies.

With the need for a scatter link established, it is necessary to evaluate this relative to other possible modes of interconnection. The scatter link has both special advantages and disadvantages and it is the present purpose to outline these rather than arrive at a conclusion as to their importance.

The large θ installation of course has a considerable portion of the cost in capital structures, whereas in the radio relay the proportionate costs of constructional engineering are higher. Scatter has a higher per station operating cost, and in general this will be higher than the commensurate microwave system. Regular replacements of such items as the high power tubes alone, may be regarded as excessively expensive.

Maintenance of the scatter system is easier, but the staff requires training and a sense of responsibility greater than that of simpler systems. Although reliability of the scatter installation is very good, a single fault in a year may outweigh all other maintenance considerations.

Logistics also tend to be more costly and difficult with the large θ installations. Strategic considerations have been discussed separately, but there are similar factors involved even in a purely commercial undertaking. Where contact is to be established over underdeveloped country, there is a real danger of disruption to on the ground facilities. Most communications companies are familiar with this problem so this need not be elaborated here.

The economic justification for the scatter link will be ultimately determined by consideration of these and other factors. For convenience the principal considerations have been outlined as follows:

1. Engineering planning
2. Survey:
   Sites
   Route
3. Construction:
   Roads
   Connecting utilities—sewage, power, etc.
   Site preparation
   Buildings
   Antenna placement
4. Equipment:
   Basic cost
   Installation
   Testing
5. Running cost:
   Administration of system
   Maintenance (Special staff. Staff training)
   Routine component replacement
   Fuel, and or power
   Supply and storage of components
   Cost of an actual breakdown
6. Deliberate disruption of service. This requires evaluation of such factors as:
   Cost of repairing damage
   Loss of revenue during period of disruption
   Cost of ultimately replacing first installed system with one of less vulnerability
7. Revenue:
   Direct on a per channel usage basis
   Indirect, arising from savings in operating costs due to the communications permitted by the link
   Losses due to lack of contact with population on intervening terrain

Most of the difficulties which mitigate against the use of scatter, do so because of the high power required. This implies that the scatter installation would become much more attractive if low power systems could become established and this in turn highlights the need for a real understanding and assessment of operational requirements rather than the present application of dogma.

Reference to Fig. 4 indicates that for the under 24-channel system a 100-watt transmitter should be extremely effective, particularly if techniques of compandors and concentrators are employed. The 100-watt system would probably be too commercially attractive to suffer over much competition from other methods.

Probable power requirements can be based on the details of Figs. 3 and 4 and the related discussions. Assuming the 40-db signal-to-noise figure, and no compandors, probable division would be

1. Under 12-channel capacity...100 watts
2. 12–60....................1,000 watts
3. 60–600..................10,000 watts

The word capacity is used deliberately here, since it permits the use of units equivalent to line concentrators, dependent on traffic activity, to provide more operative channels than the actual capacity.

For cases where lower signal/noise ratio is possible, as in good commercial or minimal military circuits, the number of channels indicated in the foregoing may be multiplied by 2–5, depending upon the limits actually set.

It must be appreciated this is purely an arbitrary division, since higher power may be called for when particularly large θ and few channels are involved. However, as already implied this rapidly becomes unjustifiable from ordinary economic considerations and ultimately one must presume these circuits are determined primarily by strategic factors.

Conclusions

This paper attempts to evaluate the scatter link not so much in terms of performance as in its compatibility with existing communication facilities, and surrounding facets of the social technology and economic structure.

Optimum performance appears to be in the 2,000-mc/sec region although choice of frequency is related to questions of availability and interference. A performance index is defined, which correlates significant system parameters. From this it is recommended, that in order to make the scatter link commercially competitive, the performance should be determined by real rather than idealized requirements. Further, it is suggested there is need for the development of equipment performing much the same function as telephone line concentrators, so that traffic can be matched to channel capacity. Compandors can also be used to advantage. Consideration of these factors indicates that links could be made much more commercially attractive. Significant results appear feasible with as little as 100-watt installations.

The particular characteristics of scatter lends itself to special applications. Strategically, the use of minimal requirements indicates that scatter angles of over 80 milliradians are feasible. In the national network, it appears the scatter link must be restricted to non-alternate backbone routes, and as such, high reliability is called for.

The normal design margins used in planning, tend to raise cost estimates and the increase due to this margin becomes disproportionate as the basic cost rises. Inviolate guarantees produce estimates which are prohibitive commercially. It is recommended that in order to produce costs which would make scatter attractive, there should be a minimal use of rigid guarantees and serious thought should be given to establishing performance parameters related to real operational requirements.

The nature of the terrain greatly affects the evaluation. It is believed there is a critical scatter angle representing the transition from microwave to scatter techniques. In many cases, it can be expected that the scatter installation can be near to a metropolis. This greatly simplifies problems of logistics, staffing, training, and maintenance, and reduces the influence of terrain on the net cost. Maintenance of the large installation requires not only greater knowledge but also an increased sense of responsibility. While this reduces the potential staff available, stations near the metropolis make greater selection likely.

The need for scatter is considered to stem basically from the socio-economic interdependence of groups of people. From this it would appear there should be a critical range of scatter angle values over which scatter is feasible. Furthermore, for the national or quasi-national networh, there is probably a critical number of channels—24 is suggested—required before the link can be justified. Links with fewer channels would be established only for special conditions determined by specific organizations, outside the national sphere. Following this argument, tentative suggestions relating transmitter power to minimal numbers of channels are suggested; it would appear from this that there is considerable utility for the 100-watt installation.

References

1. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Scatter Propagation Issue, Nov. 1955.

2. Symposium on Communications by Scatter Techniques. *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, Mar. 1956.

3. System Parameters Using Tropospheric Scatter Propagation, **H. H. Beverage, E. A. Laport, L. C. Simpson.** *RCA Review*, Princeton, N. J., vol. XVI, no. 3, Sept. 1956.

4. Dependence of the Transmission Loss in Tropospheric Radio Wave Propagation on the Angular Distance, **K. A. Norton.** *Report 3520*, National Bureau of Standards, Washington, D. C., Jan. 1955, supplement XV.

5. The Use of Angular Distance in Estimating Transmission Loss and Fading Range for Propagation Through a Turbulent Atmosphere over Irregular Terrain, **K. A. Norton, P. L. Rice, L. E. Vogler.** *Proceedings*, Institute of Radio Engineers, Oct. 1955, p. 1488.

6. A Simplified Diversity Communications System for Beyond the Horizon Links, **J. Altman, W. Sichak.** *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, Mar. 1956, p. 50.

7. Antenna to Medium Coupling Loss, **H. Staras.** *Transactions*, Professional Group on Antennas and Propagation, Institute of Radio Engineers, Apr. 1957, p. 228.

8. Transmitter RF Power Saving and Interference Reduction in the Controlled Carrier Operation of Tropospheric Scatter Communication Systems, **L. A. Ames, T. F. Rogers, E. L. Smith, L. P. Yeh.** *Symposium on Electromagnetic Interference*, Signal Corps Engineering Laboratories, Fort Monmouth, N. J., 1957.

9. Design Factors for the Optimization of Multichannel Radio Systems, **Charles A. Parry.** *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 606–20.

10. Theory of Frequency Modulation Noise, **F. Stumpers.** *Proceedings*, Institute of Radio Engineers, Sept. 1948, p. 1081.

11. Frequency Modulation Noise Characteristics, **M. G. Crosby.** *Ibid.*, Apr. 1937, p. 472.

12. International Consultative Committee on Radio, VIII Plenary Assembly, Warsaw, Poland, 1956, vols. I-III.

13. RF Bandwidth of Frequency Division Multiplex Systems Using Frequency Modulation, **R. G. Medhurst.** *Proceedings*, Institute of Radio Engineers, Feb. 1956, p. 189.

14. International Consultative Committee on Telephony, XVII Plenary Assembly, Geneva, Switzerland, Oct. 1954, vol. III.

15. The Automatic Line Concentrator and Subscriber Telephone Applications, **H. G. Evers, J. R. Wylie.** *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 6–9.

16. Load Rating Theory for Multichannel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Bell System Technical Journal*, New York, N. Y., vol. 18, 1939.

17. Application of Compandors to Telephone Circuits, **C. W. Carter, Jr., A. C. Dickieson, D. Mitchell.** *AIEE Transactions*, vol. 65, Feb. 1946, pp. 1079–86.

18. International Consultative Committee on Radio, VIII Plenary Assembly, Warsaw, 1956, vol. I.

19. Systems Engineering of Tactical Multihop Tropospheric Scatter Circuits, **A. Mack.** *Tech. Memo. M1872*, Signal Corps Engineering Laboratories, Redbank, N. J., May 1956.

20. An Extended Analysis of Echo Distortion in the FM Transmission of Frequency Division Multiplex, **R. G. Medhurst, G. F. Small.** *Proceedings*, Institute of Radio Engineers, pt. B Mar. 1956, p. 190.

The Numerical Evaluation of Expressions Involving Complete Elliptic Integrals

FREDERICK W. GROVER
MEMBER AIEE

FORMULAS involving the complete elliptic integrals K, E, and Π occur in the calculation of standards of inductance and mutual inductance, in the calculation of the magnetic field components of coils, and in the calculation of the force between coils. Other examples are found in such fields as hydrodynamics, electrostatics, and in certain applications of conformal mapping. The elliptic integrals may appear explicitly in the formulas or the computer may have to evaluate a complicated integral expression which is recognized as elliptic in nature, but in which K, E, and Π do not appear. Especial attention will here be paid to the procedure to be adopted for reducing such expressions to the standard forms.

Expressed in the Legendre nomenclature the complete elliptic integrals are

$$K=\int_0^{\frac{\pi}{2}}\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}$$

$$E=\int_0^{\frac{\pi}{2}}\sqrt{1-k^2\sin^2\varphi}\times d\varphi$$

and

$$\Pi=\int_0^{\frac{\pi}{2}}\frac{d\varphi}{(1+\eta\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}}$$

in which the modulus k lies between zero and unity, and the parameter η is unrestricted in value.

Calculations from formulas in which appear the complete integrals of first and second kinds only, namely, K and E, are usually made by interpolating their values from the tables of Legendre[1,2] or other existing tables.[3,4] In the case of a single integral, the accuracy of the result is limited only by the number of significant figures in the tabulated values, and by the accuracy of the interpolation process. However, when combinations of K and E occur in a single expression, the result obtained by substituting the tabular values of K and E, and combining them, may be markedly inferior in accuracy to the individual tabular values, due to a near cancellation of the terms. Illustrations of this are given later. In such cases the obvious way of overcoming this difficulty is by calculating K and E to a sufficient number of places to obtain the desired number of places in the result. The classic work of Legendre[1,2] employs the Landen transformation,[5] which gives the integral in terms of one with either greater or smaller modulus. (For the limiting values of $k=0$ and $k=1$, the elliptic integrals are given by elementary integrals.) It is simpler to make use of the agm (arithmetico-geometric mean) sequences of Gauss.[6] The accuracy of the values calculated by these means is limited only by the number of places on the calculating machine. However, there is an obvious advantage if the expression involving K and E can be made to depend upon the calculation of a single integral, as is illustrated in the following.

Paper 58-777, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 18, 1957; made available for printing April 1, 1958.

Frederick W. Grover is with Union College, Schenectady, N. Y.

In the case that the integral of the third kind Π enters, there is the difficulty that tables of Π do not exist, and if they did, that they would be unwieldly and inconvenient on account of the two arguments, modulus and parameter, which would be necessary. Direct calculation must be made. This has usually been accomplished by formulas in which appear both K and E and also incomplete elliptic integrals of the first and second kind, i.e., elliptic integrals where the upper limit is not $\pi/2$. The formula to be used depends upon the value of the parameter.[7] The required incomplete integrals have usually been obtained by tedious double interpolations in the tables of Legendre.[2] Much more convenient are the formulas for Π derived by King.[8] These employ trigonometric recursion formulas and the accuracy is limited only by the computing tools employed. A disadvantage for machine calculations lies in the necessity for auxiliary tables of natural trigonometric functions. Here also the formula used has to be chosen to apply to the value of the parameter.

A recognition of all these difficulties has had the effect of discouraging the use of elliptic integral formulas for computations. Series expansions of them have been preferred. For the case of inductance calculations, for example, series of hypergeometric nature are available, and also expansions in q series. Careful selection of the former is necessary to obtain sufficiently rapid convergence. The q series are very rapidly convergent but the coefficient of the general term is not known.

In 1938, Bartky[9] derived two important formulas. The first is a quadrature formula of very general scope which includes the elliptic integrals as special cases. Unfortunately, it does not yield values of Π of very great accuracy for the parameter range which is usual in physical problems, because of the rapid curvature of the curve plotted with the ordinates to be summed. However, Bartky's second formula, his equation 36 also applies to elliptic integrals and it employs recursion series which yield any desired accuracy. No auxiliary tables are required and the recursion formulas are well adapted to machine calculation. Bartky placed no especial emphasis on this formula, but it has been successfully employed in the calculation of the Bureau of Standards 3-section standard of mutual inductance.[10]

It was pointed out in 1942, in an informal memorandum by the late Dr. Frank Wenner of the Bureau of Standards, that the evaluation of complicated expressions involving complete elliptic integrals can be made to rest upon the calculation of one or the other of two type integrals, which he denoted by P and Q, or at worst, on their sum. Of these, P is Bartky's integral, his equation 36, and Q is a special case of that. This work has never been published. In the following the writer has developed and generalized this theory and has applied it to routine calculations. The illustrative examples are drawn from problems in the calculation of inductance for which series solutions have been previously made, thus giving a check.

The Integrals P and Q

Bartky's second formula may be written with small changes in nomenclature

$$P=\int_0^{\frac{\pi}{2}} \frac{s_0\cos^2\varphi+r_0p\sin^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\,\frac{d\varphi}{R} \qquad (1)$$

where $R=\sqrt{\cos^2\varphi+n_0\sin^2\varphi}=\sqrt{1-(1-n_0)\sin^2\varphi}$ which, compared with the radical $\sqrt{1-k^2\sin^2\varphi}$ of the Legendre nomenclature, shows that $n_0=1-k^2=k'^2$, k' being the complementary modulus.

The denominator $\cos^2\varphi+p\sin^2\varphi=1-(1-p)\sin^2\varphi$ is recognized as the factor $(1+\eta\sin^2\varphi)$ in the denominator of the integrand of Π, η being the parameter.

The evaluation of P depends upon the following sequences:

1. Placing $m_0=1$, and $n_0=k'$, and $c_0=k$, the gauss agm series is

$$m_1=\frac{1+n_0}{2} \qquad n_1=\sqrt{n_0} \qquad c_1=\frac{1-n_0}{2}$$

$$m_2=(m_1+n_1)/2 \qquad n_2=\sqrt{m_1n_1} \qquad c_2=\frac{m_1-n_1}{2}$$

The arithmetic means and the geometric means converge rapidly on the same limit m_∞, which is known as the arithmetrico-geometric mean.

2. Starting with $t_0=p/k'$, calculate the sequence

$$t_{i+1}=\frac{n_{m+1}(1+t_i)^2}{4m_{i+1}t_i}=\frac{n_{i+1}}{4m_{i+1}}\left(t_i+2+\frac{1}{t_i}\right)$$

and t_{i+1} converges on the limit unity.

3. Starting with r^0 $\quad r_1=\dfrac{s_0+r_0t_0}{1+t_0}$

and, in general, $\quad r_{i+1}=\dfrac{s_i+r_it_i}{1+t_i}$

the sequence $\quad s_{i+1}=\dfrac{s_i+r_i}{2}$

converges to a limiting value s_∞ and

$$P=\frac{\pi s_\infty}{2m_\infty}=s_\infty K$$

For the special case $p=1$, the integral P takes the form

$$Q=\int_0^{\frac{\pi}{2}}(z_0\cos^2\varphi+w_0\sin^2\varphi)\,\frac{d\varphi}{R} \qquad (2)$$

Starting with z_0 and w_0,

$$z_1=\frac{z_0+w_0}{2} \text{ and } m_1w_1=\frac{1}{2}(w_0+k'z_0)$$

and, in general,

$$m_{i+1}w_{i+1}=\frac{1}{2}(m_iw_i+n_iz_i).$$

The sequence

$$z_{i+1}=\frac{z_i+w_1}{2}$$

converges to the limit z_∞ and

$$Q=\frac{\pi z_\infty}{2m_\infty}$$

Integrals Involving K and E Alone

If K alone appears, the well-known result is

$$K=\frac{\pi}{2m_\infty} \qquad (3)$$

The combination $(K-E)$ may also be calculated from the m_i, n_i sequence.[11,12]

$$\frac{K-E}{K}=\frac{1}{2}(k^2+2c_1^2+4c_2^2+\ldots+2^ic_i^2+\ldots) \qquad (4)$$

It is to be noted that $(K-E)$ may also be calculated from the relation $K-E=k^2Q$, with $z_0=0$ and $w_0=1$ in the integral Q. There is little difference in the amount of labor involved in this method and the preceding one.

In the general integral expressed in terms of K and E, each with a different coefficient, it is only necessary to substitute the values

$$K=\int_0^{\frac{\pi}{2}}(\cos^2\varphi+\sin^2\varphi)\,\frac{d\varphi}{R},$$

$$K-E=\int_0^{\frac{\pi}{2}}k^2\sin^2\varphi\,\frac{d\varphi}{R}$$

$$E=\int_0^{\frac{\pi}{2}}(\cos^2\varphi+k'^2\sin^2\varphi)\,\frac{d\varphi}{R} \qquad (5)$$

and to collect terms in $\cos^2\phi$ and $\sin^2\phi$ to determine the coefficients z_0 and w_0, which apply to the form of integral Q, which is appropriate to the problem.

Example 1: The well-known Maxwell formula[13,14] for the mutual inductance of two coaxial circular filaments of radii A and a, with a distance d between their planes is

$$M=4\pi\sqrt{Aa}\left(\frac{2}{k}(K-E)-kK\right)$$

in which the modulus of K and E is $k=2\sqrt{Aa}/\sqrt{(A+a)^2+d^2}$. With the use of equations 5 the terms inside the bracket give

$$\int_0^{\frac{\pi}{2}}\left[\frac{2}{k}(k^2\sin^2\varphi)-k(\cos^2\varphi+\sin^2\varphi)\right]\frac{d\varphi}{R}=kQ$$

$z_0=-1,\ w_0=1$

so that

$$M=4\pi\sqrt{Aa}\times kQ \tag{6}$$

Using equations 3 and 4 yields

$$\frac{M}{\sqrt{Aa}}=\frac{2\pi^2}{km_\infty}[2c_1^2+4c_2^2+\ldots+2^ic_i^2+\ldots] \tag{7}$$

An equivalent equation was found by King.[15]

For two filaments of radii $a=10$, $A=90$, $d=0$, the modulus is $k=0.6$ and the complementary modulus $k'=0.8$. The details of the calculation of the agm sequence are as follows:

$m_0=1$
$n_0=0.8$
$m_1=0.9$
$n_1=0.894427191$
$m_2=0.897213596$
$n_2=0.897209268$
$m_3=0.897211432$
$n_3=0.897211432$
$m_\infty\cong 0.897211432$
$m_0w_0=1$
$n_0z_0=-0.8$
$m_1w_1=0.1$
$n_1z_1=0$
$m_2w_2=0.05$
$n_2z_2=0.049844959$
$m_3w_3=0.049922480$

$c_0=0.6$
$c_1=0.1$
$c_2=2.786404\times 10^{-3}$
$c_3=2.164\times 10^{-6}$
$z_0=-1$
$w_0=1$
$z_1=0$
$w_1=0.111111111$
$z_2=0.055555556$
$w_2=0.055728090$
$z_3=0.055641823$
$w_3=0.055641823$
$z_\infty\cong 0.055641823$

From this

$$Q=\frac{\pi}{2}\,\frac{z_\infty}{m_\infty}=0.097415133$$

$Qk=0.058449080$

and

$$\frac{M}{\sqrt{Aa}}=0.734492799$$

By equation 7, $2c_1^2+4c_2^2=0.020031056$ and, with the use of 7-place logarithms, $M/\sqrt{Aa}=0.7344930$, closely checking the result of the preceding calculation.

If this problem is calculated from the values of K and E taken from the tables of Nagaoka and Sakurai[4] there is found

$$\begin{aligned}K&=1.750754\\E&=1.418083\\\hline \text{Difference}&=0.332671\\2(K-E)&=0.665342\\k^2K&=0.630271\\\hline \text{Difference}&=0.035071\end{aligned}$$

$$\frac{M}{\sqrt{Aa}}=\frac{4\pi}{k}(0.035071)=0.734525$$

Because of the near cancellation of the two terms, this result is uncertain by 1 part in 25,000, although the tabular values are correct to seven places.

The advantage of making the calculation of Q, rather than the employment of the tables is still more strikingly apparent when the same example is solved by Maxwell's second formula[13,14] for the mutual inductance of coaxial circular filaments:

$$M=\frac{8\pi\sqrt{Aa}}{\sqrt{k_1}}[K_1-E_1] \tag{8}$$

The modulus of K_1 and E_1 is $k_1=(1-k')/(1+k')$, which is the modulus next smaller than k in the Landen transformation. For the foregoing example

$$k_1=\frac{1}{9},\ k'^2=\frac{80}{81}$$

Written in terms of the integral Q, equation 8 becomes

$$M=8\pi\sqrt{Aa}\times k_1^{3/2}Q_1 \tag{9}$$

with $z_0=0$ and $w_0=1$ in Q_1.

There is found $m_\infty=0.99690159$, $z_\infty=0.50077641$, $M/\sqrt{Aa}=0.734492799$, which checks the result by the first Maxwell equation.

The value may also be calculated by using equation 4 in equation 8. A 7-place logarithm calculation leads to the result

$$\frac{M}{\sqrt{Aa}}=\frac{12\pi^2}{m_\infty}(0.0061824247)=0.7344929$$

The calculation, using the tables, yields $K_1=1.575679$, $E_1=1.565937$, and consequently $(K_1-E_1)=0.009742$, and $M/\sqrt{Aa}=0.73453$, which is in error four units in the last place.

Inductance of a Cylindrical Current Sheet (Solenoid)

The inductance of a solenoid of radius a, axial length b, and total number of turns N, is given by Lorenz's formula.[16] The end effect of the solenoid is given by

$$\frac{L}{L_1}=\frac{4}{3\pi}\times\frac{2a}{b}\left(\frac{(2k^2-1)E+(1-k^2)K}{k^2}-1\right) \tag{10}$$

in which the modulus of the elliptic integrals is

$$k=\frac{2a}{\sqrt{4a^2+b^2}},\ \text{and}\ L_1=\frac{4\pi^2a^2N^2}{b}$$

Expressed in terms of integral Q, this becomes

$$\frac{L}{L_1}=\frac{4}{3\pi}\times\frac{2a}{b}\left[\frac{Q}{k}-1\right] \tag{11}$$

with $z_0=1$, and $w_0=2k'^2=2b^2/(4a^2+b^2)$.

Example 2: For a single-layer coil, wound on a marble cylinder, to serve as an inductance standard at the Bureau of Standards, the dimensions of the equivalent current sheet are $2a=54.1724$ centimeters, $b=30.5510$ centimeters, the total number of turns being $N=440$.

These dimensions yield the modulus $k=0.871032075$, and the complementary modulus $k'=0.491226143$. The convergent of the agm series $1,k'$ is $m_\infty=m_5=0.723071158$ and, calculating $z_\infty\cong z_5=0.696489721$, there follows $Q=1.51305094$, $Q/k=1.73707833$, $8a/3\pi b=0.752560656$. Therefore $L/L_1=0.554696149$, and $L_1=0.183542152$, so that $L=0.101810125$ henry. This value checks closely with that found in previous work.[17]

Equation 10 gives the inductance as the difference of two terms which correspond to two values of axial distance. This is inherent in the nature of the formula.

Equation 4 could be used in the calculation but this would reduce equation 10 to an unfavorable form; in fact, the elegant expression for coaxial filaments seems to be an exception. In general, for routine computations, expressions involving integral Q are preferable.

This reduces calculation to a simple procedure. The sequence (m_i, n_i) is first calculated from the numbers 1, k'. Then, starting from the given values of z_0 and w_0, z_1 is their arithmetic mean. The arithmetic mean of the calculated values m_0w_0 and n_0z_0 is m_1w_1, and dividing this by m_1 yields w_1. That is, each convergent z_{i+1} is the mean of z_i and w_i, and each product $w_{i+1}\times m_{i+1}$ is the mean of w_im_i and z_in_i. The quantities m_i and n_i are available from the sequence formed from $m_0=1$ and $n_0=k'$.

Other important applications of integral Q are, for the components of the magnetic field intensity, due to a current of I amperes, flowing in a circular filament of radius a. At a point P situated at an axial distance x from the plane of the filament, and at a distance y from the axis, Russell's formulas[18,19] become as follows:

The axial component:

$$H_x=\frac{2NI}{10a\rho_1}\left(\frac{2(1-y/a)}{\rho_2^2}\times E+(K-E)\right)=\frac{4NI}{10a\rho_1}\times Q \tag{12}$$

with $z_0=(1-y/a)/\rho_2^2$, and $w_0=(1+y/a)/\rho_1^2$ in Q.

$$\rho_1^2=(1+y/a)^2+\frac{x^2}{a^2} \qquad \rho_2^2=(1-y/a)^2+\frac{x^2}{a^2}$$

The radial component:

$$H_r=\frac{2NI}{10a\rho_1}\left[\frac{2x/a}{\rho_2^2}-\frac{x}{y}(K-E)\right]=\frac{2NI}{10a\rho_1}\times Q \tag{13}$$

with

$$z_0=\frac{2x/a}{\rho_2^2} \text{ and } w_0=-\frac{2x/a}{\rho_1^2}$$

in Q in equation 13. The modulus of the elliptic integrals is

$$k=\frac{2\sqrt{y/a}}{\rho_1} \text{ and } k'=\rho_2/\rho_1.$$

Maxwell's formula for the mechanical force in dynes,[13,20] exerted between two coaxial filaments of radii a and A, each carrying a current of one ampere, when the distance between their planes is d, may be written

$$F=\frac{\pi r_1 d}{r_2^2}[(4-2k^2)E-4k'^2K]$$

$$=\frac{2\pi r_1 d k^2}{r_2^2}\times Q \tag{14}$$

with $z_0=1$, $w_0=-k'^2$ in Q. The modulus is $k=2\sqrt{Aa}/r_1$, and $r_1^2=(A+a)^2+d^2$ and $r_2^2=(A-a)^2+d^2$.

In Appendix I are given the expressions for certain elliptic integrals in terms of integral Q. These may be useful in certain complicated expressions. The reduced expression in terms of a single integral Q is obtained by combining the terms in $\cos^2\phi$ in the integrand to obtain the value of z_0, and by combining the terms in $\sin^2\phi$, the value of w_0 results.

Complete Integral of the Third Kind

The complete elliptic integral of the third kind Π, may be calculated directly by means of the integral P, with $s_0=1$, $r_0 p=1$, $p=1+\eta$, and therefore $r_0=1/1+\eta$. The parameter η in P is unrestricted in value, except for the case that η lies between -1 and $-\infty$, which requires special treatment.

For this case, placing $\eta=-1/\sin^2\theta$, the value of $\sin^2\theta$, corresponding to the given value of η, is calculated, and then the required integral is found from the relation[21]

$$\Pi\left(-\frac{1}{\sin^2\theta}, k\right)=K(k)-\Pi(-k^2\sin^2\theta, k) \tag{15}$$

This makes the calculation of the required integral depend upon the value of the integral P with the parameter $-k^2\sin^2\theta$ substituted for η.

For most physical applications, the parameter is negative and lies between the limits $-k^2$ and -1. It is therefore customary to write $\eta=-c^2$, so that $p=1-c^2=c'^2$ and $1-p=c^2$.

In most problems Π does not appear alone, but in the combination $(\Pi-K)$ for which P takes the form

$$(\Pi-K)=\int_0^{\frac{\pi}{2}}\frac{c^2\sin^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\times\frac{d\varphi}{R} \tag{16}$$

which is integral P with $s_0=0$, $p=1-c^2$, and $r_0=(1-p)/p=c^2/c'^2$.

A special case of the elliptic integral of third kind is that where the parameter $-c^2=-k^2$. This may, of course, be treated as an integral P with the coefficient $p=k'^2$. However, it is possible to show that this special integral may be expressed in terms of K and E only, and consequently may be calculated also as an integral Q, with appropriate values of z_0 and w_0. In Appendix I are noted certain integrals of this class, together with the coefficients which apply, both for use with integral P and with integral Q.

Integrals Involving (Π−K) and K Alone

Writing

$$I=\alpha(\Pi-K)+\gamma K \tag{17}$$

and substituting equation 16 for equation $(\Pi-K)$ and

$$\int_0^{\frac{\pi}{2}}\frac{d\phi}{R} \text{ for } K$$

and reducing to a common denominator yields

$$I=\int_0^{\frac{\pi}{2}}\frac{\gamma\cos^2\varphi+(\alpha c^2+p\gamma)\sin^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\times\frac{d\varphi}{R} \tag{18}$$

which is integral P with

$$s_0=\gamma,\ r_0=\frac{\alpha(1-p)}{p}+\gamma=\alpha\frac{c^2}{c'^2}+\gamma$$

As an example, the axial component of the magnetic field strength at a point situated in the end plane of a solenoid is [22,23]

$$H_{zs}=\frac{2N}{r_1}(1+c')[K\pm c'(\Pi-K)] \tag{19}$$

where

a = radius of the solenoid
x = length of the solenoid
y = distance of the point from the axis
N = number of turns

$$r_1=\sqrt{(a+y)^2+x^2} \qquad k=\frac{2\sqrt{ay}}{r_1}$$

$$c=\frac{2\sqrt{ay}}{(a+y)} \qquad c'^2=\frac{(a-y)^2}{(a+y)^2}$$

The positive sign is to be used for points with $y<a$, and the negative for points for which $y>a$. Equation 19 becomes

$$H_{zs}=\frac{4aN}{r_1(a+y)}\times P \tag{20}$$

in which $s_0=1$, and $r_0=1/c'$ for $y<a$, and $r_0=-1/c'$ for $y>a$.

Integrals which May Be Expressed by (P+Q)

Making use of the defining equations for P and Q, and reducing to a common denominator, there results the general expression

$$P+Q=\int_0^{\frac{\pi}{2}}\frac{(s_0+z_0)\cos^2\varphi+p(r_0+w_0)\sin^2\varphi+(1-p)(w_0-z_0)\sin^2\varphi\cos^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\times\frac{d\varphi}{R} \tag{21}$$

This expression will be used here to treat the general integral involving K, E, and Π. First, however, will be considered an especially simple case.

Jones' Equation

Jones' equation[24] is for the mutual inductance of a solenoid and a circular filament in its end plane:

$$M=2\pi N(A+a)ck\left(\frac{K-E}{k^2}-\frac{c'^2}{c^2}(\Pi-K)\right) \tag{22}$$

where

N = number of turns in an axial length x of solenoid
a,A = the radii of solenoid and filament

$$k^2=\frac{4Aa}{(A+a)^2+x^2},\ c^2=\frac{4Aa}{(A+a)^2},\ c'^2=\frac{(A-a)^2}{(A+a)^2}$$

Jones' formula was first used by Campbell,[22] page 428, at the National Physical Laboratory to calculate the constant of an absolute standard of mutual inductance. In its use at the Bureau of Standards in determining the constant of the 3-section standard of mutual inductance,[10] its evaluation has been effected by calculating the first term by equation 4, while the second term was calculated from the Bartky integral (integral P here). This procedure gives the constant with any desired accuracy. The two terms have opposite signs, but since one term was two to five times the other, this was unimportant.

The Jones formula may also be treated as a special case of the sum $(P+Q)$, making use of the general equation 21. However, the coefficients of the two terms are so related that a still simpler equation results.

Substituting for $(K-E)$ and $(\Pi-K)$

from equations 5 and 16, there is found

$$\frac{K-E}{k^2}-\frac{c'^2}{c^2}(\Pi-K)$$

$$=\int_0^{\frac{\pi}{2}}\frac{(1-p)\sin^2\varphi\cos^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\times\frac{d\varphi}{R}\qquad(23)$$

and, comparing this with general equation 21, note that, if the relations hold,

$$s_0+z_0=0 \text{ and } r_0+w_0=0\qquad(24)$$

then the Jones formula becomes

$$M=\frac{2\pi N(A+a)ck(P+Q)}{w_0-z_0}$$

$$=\frac{2\pi N(A+a)ck}{w_0-z_0}\frac{\pi}{2}\frac{(s_\infty+z_\infty)}{m_\infty}$$

$$=\frac{\pi^2N(A+a)ck}{s_0-r_0}\left(\frac{s_\infty+z_\infty}{m_\infty}\right)\qquad(25)$$

This equation, which was derived by Dr. Wenner, gives great flexibility of solution. As long as the conditions, equation 24, are satisfied, s_0 and r_0 may be selected at will and, furthermore, check solutions may be made using different assumed values of these quantities, without the necessity of recomputing the (m_i,n_i) and t_i sequences.

Example 3. Campbell's example, with $a=10$, $A=14.5$, $x=10$. From the foregoing dimensions, there are found

$k'=0.414396409$, $c=0.98298732$

$k=0.91009649$, $p=c'^2=0.0337359433$

and the $(m_i\,n_i)$ and t_i sequences are

$m_0=1$	$t_0=p/k'$
$n_0=0.414396409$	$=0.0814098350$
$m_1=0.707198204$	$t_1=3.268967504$
$n_1=0.643736288$	
$m_2=0.675467246$	$t_2=1.392179960$
$n_2=0.674721532$	
$m_3=0.675094389$	$t_3=1.027619144$
$n_3=0.675094172$	
$m_4=0.675094280$	$t_4=1.000185579$
$n_4=0.675094280$	
$m_\infty\cong0.675094280$	$t_5=1.000000008$

The calculation of $(P+Q)$ with different assumptions follows.

1	2	3
$s_0=1,\ r_0=-2$	$s_0=1,\ r_0=-1$	$\frac{c'^2}{c^2}(\Pi-K)$ by P
$z_0=-1,\ w_0=2$	$z_0=-1,\ w_0=1$	$s_0=0,\ r_0=1$
$s_0-r_0=3$	$s_0-r_0=2$	$\frac{K-E}{k^2}$ by equation 4
$s_\infty\cong s_5=0.320355180$	$s_\infty=0.5469034$	
$z_\infty\cong z_5=0.817818682$	$s_\infty=0.2118791$	$\frac{K-E}{k^2}=\frac{\pi}{2m_\infty}(0.6059396)$
sum $=1.138173862$	sum $=0.7587825$	$\frac{c'^2}{c^2}(\Pi-K)=\frac{\pi}{2m_\infty}(0.2265482)$
$\frac{1}{3}$sum $=0.379391287$	$\frac{1}{2}$sum $=0.37939125$	Difference $=\frac{\pi}{2m_\infty}(0.3793914)$
$\frac{P+Q}{s_0-r_0}=0.882760314$	$\frac{P+Q}{s_0-r_0}=0.8827603$	$=0.8827607$

The first column, which is a machine calculation, leads, using the number of turns in Campbell's example,[22] to the value $M=24313.8705$ abhenrys per turn on the secondary, checking Campbell's value closely. The calculations in the second and third assumptions were made by 7-place logarithms.

The General Case

The general equation involving complete elliptic integrals of all three kinds may be conveniently written

$$I=\alpha(\Pi-K)+\beta(K-E)+\gamma K\qquad(26)$$

and, making use of equations 16, and 5, yields the integral

$$I=\int_0^{\frac{\pi}{2}}\frac{\gamma\cos^2\varphi+(\alpha c^2+\beta k^2p+p\gamma)\times\sin^2\varphi+(1-p)\beta k^2\sin^2\varphi\cos^2\varphi}{\cos^2\varphi+p\sin^2\varphi}\times\frac{d\varphi}{R}\qquad(27)$$

Comparing this integral with that of equation 21, shows that the general integral 26 may be represented by $(P+Q)$, if the following conditions are satisfied:

$$s_0+z_0=\gamma$$

$$p(r_0+w_0)=\alpha c^2+\beta k^2p+p\gamma$$

$$(1-p)(w_0-z_0)=\beta k^2(1-p)$$

or

$$s_0+z_0=\gamma$$

$$r_0+w_0=\alpha\frac{c^2}{c'^2}+\beta k^2+\gamma\qquad(28)$$

$$w_0-z_0=\beta k^2$$

The choice of one of the coefficients is arbitrary, provided these relations between the coefficients are satisfied.

Example 4. Mutual inductance of coaxial solenoids: This case has received considerable attention, and the work of Nagaoka,[25] Terezawa,[26] Olshausen[27] and Cohen[28] may be cited. If the radii are a and A, the lengths $2l$ and $2l_1$, and the distance between their centers δ, then the mutual inductance is of the form

$$M=I_1-I_2-I_3+I_4$$

in which the terms are the values of the basic integral, given by Cohen:

$$I=\frac{32\pi nn_1\sqrt{(A+a)^2+\delta^2}}{(A+a)^2}\times$$

$$\int_0^{\frac{\pi}{2}}\frac{\sin^2\varphi\cos^2\varphi(1-k^2\sin^2\varphi)d\varphi}{(1+\eta\sin^2\varphi)\sqrt{1-k^2\sin^2\varphi}}\qquad(29)$$

with the four distances $(d+l+l_1)$, $(d+l-l_1)$, $(d-l+l_1)$, $(d-l-l_1)$ respectively, substituted for δ. In equation 29 n and n_1 are the winding densities. For the case where the solenoids are concentric, as well as coaxial, $I_4=I_1$ and $I_3=I_2$, so that $M=2(I_1-I_2)$.

The numerator of equation 29 is $\sin^2\phi-(1+k^2)\sin^4\phi+k^2\sin^6\phi$ and, expanding the expression by the method given in the Appendix II, there results

$$I=\frac{8\pi nn_1}{3}\times Aa\sqrt{(A+a)^2+\delta^2}\times(P+Q)\qquad(30)$$

in which

$$\alpha=-\frac{3(c^2-k^2)(1-k^2)}{c^4},\ \beta=\frac{c^2(1+k^2)-3k^2}{c^2k^2},$$

$$\gamma=1$$

subject to the conditions $s_0+z_0=1$, $r_0+w_0=-k'^2$, $w_0-z_0=1+k^2-3k^2/c^2$ or, expressed in terms of z_0

$$s_0=1-z_0$$

$$w_0=z_0+(1+k^2)-\frac{3k^2}{c^2}\qquad(31)$$

$$r_0=-z_0-2+\frac{3k^2}{c^2}$$

In the example proposed by Cohen[28] and also treated by Nagaoka,[25] the concentric solenoids have $a=10$, $A=15$, $2l=20$, $2l_1=200$, so that for I_1, $\delta=110$, and for I_2, $\delta=90$.

For the calculation of I_1 the condition equations are

$$s_0=1-z_0$$

$$r_0=-z_0-1.852652259$$

$$w_0=z_0+0.899803537$$

and also

$k^2=0.047151277$	$t_0=0.040977908$
$k'^2=0.952848723$	$t_1=6.610635676$
$-\eta=c^2=0.96$	$t_2=2.190476763$
$p=c'^2=0.04$	$t_3=1.161749596$
$m_\infty\cong m_3=0.988033841$	$t_4=1.005630071$
	$t_5=1.000007880$

$$k^2=\frac{4Aa}{(A+a)^2+\delta^2}$$

$$c^2=\frac{4Aa}{(A+a)^2}$$

Making the calculation for two assumed values of z_0,

$z_0=-0.5$

$s_0=1.5$

$r_0=-1.352652259$

$w_0=0.399803537$

$s_\infty \cong s_7=1.019744972$

$z_\infty \cong z_4=-0.047382082$

$s_\infty+z_\infty=0.972362890$

$z_0=0.5$

$s_0=0.5$

$r_0=-2.352652259$

$w_0=1.399803537$

$s_\infty \cong s_6=0.019744972$

$z_\infty \cong z_4=0.952617918$

$s_\infty+z_\infty=0.972362895$

Using the mean of these values and those for $\delta=90$ also, yields

for $\delta=110$, $\sqrt{12725}\,(P'+Q')=174.3834753$

for $\delta=\;90$, $\sqrt{8725}\;\;(P''+Q'')=143.3159327$

Difference $=\;31.0675426$

These lead to the value $M=6213.50852$ $(4\pi nn_1)$, which confirms the value 6213.52 by Nagaoka's formula and 6213.510 by the Searle and Airey[29] formula, both computed with the use of 7-place logarithms.

An examination of the foregoing values of s_∞ and z_∞ shows that a change Δz_0 in the assumed value of z_0 results in an equal change of Δz_0 in the end value z_∞, and a change of $-\Delta z_0$ in the value of s_∞. This is easily shown to be a general truth, and makes it readily possible to choose z_0 so that z_∞ shall be the larger part of the sum $(z_\infty+z_\infty)$. This has the advantage that, when the values of t_i converge only slowly toward their limit of unity, as in this example, the convergence of the sequence z_i is better than that of s_i. A second calculation of z_∞ and s_∞, with an amended value of the assumed value z_0, on which the calculation is based, adds little to the labor involved, since the value of m_∞ and the values of t_i are unchanged. Also, at the each step of the calculation of the new sequences, z_i and s_i, the values should differ from the previously calculated values by the difference Δz_0 between the two assumed values of z_0. This allows errors of calculation to be easily located.

An examination of the foregoing calculated values shows that if z_0 had been assumed to be $z_0=0.519744977$, and therefore $s_0=1-z_0=0.480255023$, then s_∞ would result equal to zero, and the calculation would have depended upon Q alone. This is in line with the fact that $(\Pi-K)$ may be expressed in terms of K, the factor being $(1-p)/p\times s_\infty''$, where s_∞'' is the convergent of the sequence with $s_0=0$ and $r_0=1$. This procedure is found to confirm the critical value of z_0 just cited, but the labor is not very different, since the sequence s_i'' must be calculated. In addition, by this procedure the advantage of a check calculation is lost.

The Jones' formula may be considered as a special case of the general equation 26 with $\gamma=0$. Taking account of the values of α and β, the condition equations found are $s_0=-z_0$, $w_0=-r_0=1+z_0$, which are included in the more general equations 24.

Appendix I. Some Integrals Which Depend Upon Q for Their Evaluation

With $k=$ the modulus, and $k'=\sqrt{1-k^2}=$ the complementary modulus,

$$R=\sqrt{1-k^2\sin^2\varphi}=\sqrt{\cos^2\varphi+k'^2\sin^2\varphi}$$

$$\int_0^{\frac{\pi}{2}}\sqrt{1-k^2\sin^2\varphi}\times d\varphi=E=Q \text{ with } z_0=1,\ w_0=k'^2$$

$$\int_0^{\frac{\pi}{2}}\sin^2\varphi\times\frac{d\varphi}{R}=\frac{1}{k^2}(K-E)=Q \text{ with } z_0=0,\ w_0=1$$

$$\int_0^{\frac{\pi}{2}}\cos^2\varphi\times\frac{d\varphi}{R}=\frac{1}{k^2}[E-k'^2K]=Q \text{ with } z_0=1,\ w_0=0$$

$$\int_0^{\frac{\pi}{2}}\frac{d\varphi}{R^3}=\frac{1}{k'^2}E=\frac{Q}{k'^2} \text{ with } z_0=1,\ w_0=k'^2$$

$$\int_0^{\frac{\pi}{2}}\frac{\cos^2\varphi}{R^3}\,d\varphi=\frac{1}{k^2}(K-E)=Q \text{ with } z_0=0,\ w_0=1$$

$$\int_0^{\frac{\pi}{2}}\frac{\sin^2\varphi}{R^3}\,d\varphi=\frac{1}{k^2k'^2}[E-k'^2K]=\frac{Q}{k'^2} \text{ with } z_0=1,\ w_0=0$$

$$\int_0^{\frac{\pi}{2}}R^3d\varphi=\frac{1}{3}[2(1+k'^2)E-k'^2K]=\frac{Q}{3} \text{ with } z_0=2+k'^2,\ w_0=k'^2(1+2k'^2)$$

$$\int_0^{\frac{\pi}{2}}R\sin^2\varphi d\varphi=\frac{(2k^2-1)E+k'^2K}{3k^2}=\frac{Q}{3} \text{ with } z_0=1,\ w_0=2k'^2$$

$$\int_0^{\frac{\pi}{2}}R\cos^2\varphi d\varphi=\frac{(1+k^2)E-k'^2K}{3k^2}=\frac{Q}{3} \text{ with } z_0=2,\ w_0=k'^2$$

$$\int_0^{\frac{\pi}{2}}\sin^4\varphi\times\frac{d\varphi}{R}=\frac{(2+k^2)K-2(1+k^2)E}{3k^4}=\frac{Q}{3k^2} \text{ with } z_0=-1,\ w_0=1+2k^2$$

$$\int_0^{\frac{\pi}{2}}\cos^4\varphi\times\frac{d\varphi}{R}=\frac{(2-3k^2)k'^2K-2(1-2k^2)E}{3k^4}=\frac{Q}{3k^2} \text{ with } z_0=3k^2-1,\ w_0=k'^2$$

$$\int_0^{\frac{\pi}{2}}\frac{\sin^2\varphi\cos^2\varphi}{R^3}\times d\varphi=\frac{(1+k'^2)K-2E}{k^4}=\frac{Q}{k^2} \text{ with } z_0=-1,\ \ w_0=1$$

$$\int_0^{\frac{\pi}{2}}\sin^2\varphi\cos^2\varphi\times\frac{d\varphi}{R}=\frac{(2-k^2)E-2k'^2K}{3k^4}=\frac{Q}{3k^2} \text{ with } z_0=1,\ w_0=-k'^2$$

$$\int_0^{\frac{\pi}{2}}\sin^6\varphi\times\frac{d\varphi}{R}=\frac{1}{15k^6}[(8+3k^2+4k^4)K-(8+7k^2+8k^4)E]=\frac{Q}{15k^4} \text{ with } z_0=-4(1+k^2),\ w_0=4+3k^2+8k^4$$

$$\int_0^{\frac{\pi}{2}}\cos^6\varphi\times\frac{d\varphi}{R}=\frac{1}{15k^6}[k'^2(-8+19k^2-15k^4)K+(8-23k^2+23k^4)E]=\frac{Q}{15k^4} \text{ with } z_0=4-11k^2+15k^4,\ w_0=-4k'^2(1-2k^2)$$

Appendix II

For the reduction of the elliptic integral

$$\int_0^{\frac{\pi}{2}}\sum_n\left[\frac{\sin^{2n}\varphi}{(1+\eta\sin^2\varphi)^r}\times\frac{d\varphi}{R}\right]$$

the recursion formula (i) on page 60 of reference 30 will be found useful. For complete integrals this reads

$$-2\mu AV_{\mu-1}+(2\mu+1)BV_\mu-(2\mu+2)CV_{\mu+1}+(2\mu+3)k^2V_{\mu+2}=0$$

in which

$$V_\mu=\int_0^{\frac{\pi}{2}}(m+\sin^2\varphi)^\mu\times\frac{d\varphi}{R}$$

$$A=m(1+m)(1+k^2m)$$

$$B=1+2m(1+k^2)+3k^2m^2$$

$$C=1+k^2+3k^2m$$

For example, with $m=1/\eta$ and $\mu=-1$, $2AV_{-2}=BV_{-1}-k^2V_1$, where the integrals are

$$V_{-2}=\int_0^{\frac{\pi}{2}}\frac{\eta^2}{(1+\eta\sin^2\varphi)^2}\times\frac{d\varphi}{R};$$

$$V_{-1}=\int_0^{\frac{\pi}{2}}\frac{\eta}{(1+\eta\sin^2\varphi)}\times\frac{d\varphi}{R}=\eta\Pi$$

$$V_1=\int_0^{\frac{\pi}{2}}(1+\eta\sin^2\varphi)\times\frac{d\varphi}{R}=\frac{K}{\eta}+\frac{(K-E)}{k^2}$$

With $\eta=0$, $m=0$, $\mu=n$, the general integral $\int_0^{\frac{\pi}{2}}\sin^{2n}\varphi\times d\varphi/R$ may be evaluated.

The evaluation of certain integrals in the text follows.

$$\int_0^{\frac{\pi}{2}}\frac{\sin^2\varphi}{(1+\eta\sin^2\varphi)}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta}\int_0^{\frac{\pi}{2}}\frac{(\eta\sin^2\varphi+1-1)}{1+\eta\sin^2\varphi}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta}(K-\Pi)$$

$$\int_0^{\frac{\pi}{2}}\frac{\sin^4\varphi}{(1+\eta\sin^2\varphi)}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^2}\int_0^{\frac{\pi}{2}}\frac{(\eta^2\sin^4\varphi-1+1)}{1+\eta\sin^2\varphi}\times\frac{d\varphi}{R}$$

$$=-\frac{1}{\eta^2}\int_0^{\frac{\pi}{2}}(1-\eta\sin^2\varphi)\times\frac{d\varphi}{R}+\frac{1}{\eta^2}\int_0^{\frac{\pi}{2}}\frac{1}{1+\eta\sin^2\varphi}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^2}(\Pi-K)+\frac{K-E}{k^2\eta}$$

$$\int_0^{\frac{\pi}{2}}\frac{\sin^6\varphi}{(1+\eta\sin^2\varphi)}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^3}\int_0^{\frac{\pi}{2}}\frac{(\eta^3\sin^6\varphi+1-1)}{1+\eta\sin^2\varphi}\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^3}\int_0^{\frac{\pi}{2}}\left[1-\eta\sin^2\varphi+\eta^2\sin^4\varphi-\frac{1}{1+\eta\sin^2\varphi}\right]\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^3}(K-\Pi)-\frac{(K-E)}{k^2\eta^2}+\frac{1}{\eta}\int_0^{\frac{\pi}{2}}\sin^4\varphi\times\frac{d\varphi}{R}$$

$$=\frac{1}{\eta^3}(K-\Pi)-\frac{(K-E)}{k^2\eta^2}+\frac{1}{\eta}\left[\frac{2(1+k^2)(K-E)-k^2K}{3k^4}\right]$$

References

1. Traite des Fonctions Elliptiques (book), **A. M. Legendre.** 1825.

2. A. M. Legendres Tafeln der Elliptischen Normalintegrale (book), **Fritz Emde, Konrad Wittwer,** editors. Stuttgart, Germany, 1931.

3. Formulas and Tables for the Calculation of Mutual and Self-Induction, **E. B. Rosa, F. W. Grover.** *Bulletin,* National Bureau of Standards, Washington, D. C., vol. 8, 1911, p. 1; also *Scientific Paper no. 169,* National Bureau of Standards, 1912, pp. 202–12, Tables XII and XIII.

4. **H. Nagaoka, S. Sakurai.** *Scientific Papers,* Institute of Physical and Chemical Research, Tokyo, Japan, Dec. 1922, Table I.

5. **Landen.** *Philosophical Transactions,* Royal Society of London, London, England, vol. 65, 1775, p. 283.

6. On the Direct Calculation of Elliptic Functions and Integrals (book), **L. V. King.** Cambridge University Press, Cambridge, England, 1924, pp. 6, 8.

7. Elliptic Functions (book), **Cayley.** George Hall and Sons, London, England, 1895, p. 139.

8. See reference 6, pp. 19–30.

9. Numerical Calculation of a Generalized Elliptic Integral, **Walter Bartky.** *Reviews of Modern Physics,* New York, N. Y., vol. 10, 1938, p. 264.

10. An Absolute Measurement of Resistance by the Wenner Method, **Thomas, Peterson, Cooter, Kotter.** *Journal of Research,* National Bureau of Standards, vol. 43, 1949, p. 312.

11. **L. V. King.** *Philosophical Magazine,* London, England, vol. 15, ser. 7, 1933, p. 1101.

12. **F. W. Grover.** *Ibid.,* p. 1117.

13. **John Clark Maxwell.** *Electricity and Magnetism,* Oxford University Press, London, England, vol. II, art. 701.

14. See reference 3, pp. 6–7.

15. See reference 11, p. 1102, equation 18.

16. **Lorenz.** *Wiedemann's Annalen,* Leipzig, Germany, vol. 7, 1879, p. 161; also reference 3, p. 118, formula 72.

17. See reference 3, p. 129, example 59.

18. Alternating Currents, Vol. I (book), **A. Russell.** Cambridge University Press, Cambridge, England, 1914, p. 98.

19. Formulas and Tables for the Calculation of the Magnetic Field Components of Circular Filaments and Solenoids, **Frederick W. Grover.** *AIEE Transactions,* vol. 68, pt. I, 1949, p. 666, formulas 3 and 4.

20. **Rayleigh.** *Philosophical Transactions,* vol. 175, 1884, pp. 411–60.

21. See reference 7, p. 123.

22. On a Standard of Mutual Inductance, **A. Campbell.** *Proceedings,* Royal Society of London, vol. 79, p. 434, formula 4.

23. See reference 19, formulas 22 and 23.

24. Calculation of the Coefficient of Mutual Induction of a Circle and a Coaxial Helix, **J. V. Jones.** *Proceedings,* Royal Society of London, vol. 63, 1898, p. 192; also reference 3, p. 109.

25. **H. Nagaoka.** *Journal,* College of Science, Imperial University of Tokyo, Tokyo, Japan, vol. 27, 1909, art. 6; also reference 3, pp. 64–67; example 43, p. 91.

26. **K. Terezawa.** *Tokyo Mathematical Journal* Tokyo, Japan, vol. 10, 1916, p. 73.

27. **G. Olshausen.** *The Physical Review,* New York, N. Y., vol. 31, 1910, p. 617.

28. **Louis Cohen.** *Bulletin,* National Bureau of Standards, vol. 3, 1907, p. 295; also reference 3, p. 64.

29. **Searle, Airey.** *The Electrician,* London, England, vol. 56, 1905, p. 318; also reference 3, p. 61.

30. Elliptic Integrals (book), **H. Hancock.** John Wiley & Sons, Inc., New York, N. Y., 1917, chap. IV.

31. Elliptic Functions (book), **Baker.** John Wiley & Sons, Inc., 1895, chap. IV.

Vol. 7
Pt. 1
1958

Core-Reset Functions in Magnetic-Amplifier Analysis

Part I—A Core-Reset Function

G. C. FETH
ASSOCIATE MEMBER AIEE

THE OPERATION of a magnetic amplifier is described by the differential equations of its physical system. In any analysis of amplifier operation the circuit equations resulting from the application of Kirchhoff's laws possess terms containing the voltages which are induced electromagnetically in the core windings by changing core flux in accordance with Faraday's law, and terms containing the various currents which flow in the windings. In general, a knowledge of the core properties relating instantaneous values of the rate of change of core flux ϕ' and of net mmf (magnetomotive force) F of the currents in the windings is an essential part of a complete description of the physical system.

The common, statically measured magnetic quantities are inadequate to describe core behavior under dynamic conditions of changing flux. Mathematical derivations of the over-all relationship between mmf and rate of change of core flux have been hindered by the excessively complicated mathematical problems encountered and by an inaccurate knowledge of the values of the associated physical constants. Qualitative reference to dynamic effects in magnetic cores is often made in terms of the dependence of the width of the dynamic hysteresis loops upon frequency; however, such expression of these effects tends to obscure the instantaneous relationship between the magnetic variables which describe the core's behavior.

It is the purpose of this paper to develop, consistently with the present understanding of flux reversal in magnetic tapes, a general analytical relationship between instantaneous values of the magnetic variables describing the behavior of tape cores of grain-oriented nickel-iron alloy. The resulting analytical description of core behavior will be used in a companion paper to calculate the performance of various magnetic amplifiers; one such example will be included in this paper as a check on the validity of the reset function.

Variables Describing Magnetic State of Core

The dynamic processes of magnetization in tape cores of grain-oriented nickel-iron alloys are described in reference 1 by means of certain theoretical models. Flux changes are ascribed to the movement of the domain walls separating regions of negative from regions of positive saturation flux density. The application of a sufficiently large reverse mmf to a saturated core brings about the formation of small regions of reversed saturation flux density along the edges of the tape cross section and causes the domain walls thus produced to enlarge and move through the tape; the number of nuclei of flux reversal so formed depends largely upon the magnitude of the applied mmf. The motion of the domain walls is hindered by eddy currents, which flow in conducting paths encircling regions of changing flux density; by spin-relaxation phenomena, whose effect upon wall motion is analogous to viscous damping; and by other effects. On the basis of experimental evidence it was concluded that time derivatives of flux other than the first derivative have no appreciable effects on the magnetization process at power frequencies. In short, it appears that dynamic effects of magnetization are dependent both upon the magnitude of F, which affects the number of places where flux reversal initiates and provides the motivation for domain wall motion, and upon ϕ', which produces eddy current and spin-relaxation effects retarding the flux reversal.

It is quickly recognized that the variables F and ϕ' alone are not sufficient to describe core reset; e.g., for reset at constant mmf many values of ϕ' may occur, implying that F and ϕ' are not uniquely related. From the model it appears quite reasonable that the mmf required to produce a given ϕ' may well depend upon both the location and shape of the domain walls. For instance, at a given value of ϕ', implying a given rate of growth of regions of reversed magnetization, the effects of induced eddy currents may be more or less pronounced, depending upon the eddy current paths, which in turn are affected by the geometry of the domain walls and their proximity to the tape surface and to each other. Generally, the configuration of the domain walls is the essential physical feature which makes the "magnetic history" of a material so important in determining its future behavior. In this light, it would appear that any attempt to describe core behavior must consider the point-by-point flux density distribution throughout the material. However, it remains to be seen whether the accurate knowledge of the domain wall configuration is truly indispensable for fairly modest, practical purposes.

For these purposes, it is desirable that the history of the core be expressed by some clearly defined and easily measured magnetic quantity. The flux ϕ existing in the core at any given instant may well be such a quantity; its usefulness in specifying the saturation flux level is apparent. Identity of domain wall configurations implies identity of fluxes. The reverse is not necessarily true, since equality of fluxes does not imply identity of wall configurations, but it seems not unreasonable that, starting from identical wall configurations (as is the case in reset processes initiated from pronounced saturation) and arriving at equal fluxes, the wall configurations may not differ significantly even for somewhat different reset processes. Hence, core flux may well be accepted tentatively as a quantity embodying the magnetic history of the core for flux-reset processes initiating in saturation; this leads to the simplification that the magnetic state of the core during reset may be expressed in terms of the variables F, ϕ', and ϕ.

Certain restrictions on the use of this simplification are necessary. In flux-reversal processes initiated from flux levels other than saturation, ϕ alone is not likely to embody adequately the implications of core history,[2] which explains the restriction to processes initiating from saturation. If during a reset process the mag-

---

Paper 57-95, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 5, 1956; made available for printing November 21, 1956.

G. C. Feth is with the General Electric Company, Schenectady, N. Y.

This work has been performed in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology. The work has been supported both by the General Electric Educational and Charitable Fund through a Coffin Fellowship and by the United States Office of Naval Research. The author wishes to acknowledge the valuable guidance of his thesis supervisor, Prof. L. A. Finzi of Carnegie Institute of Technology.

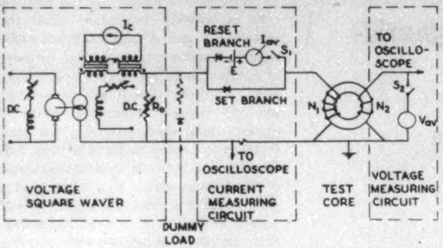


Fig. 1. Test circuit for F_{av} versus ϕ' test

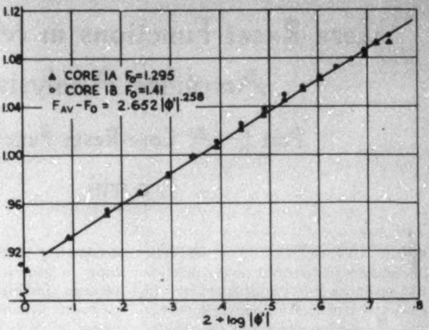


Fig. 3 (right). Log-log plot of F_{av} versus ϕ' test for cores 1A and 1B

nitude of F increases appreciably, additional nuclei of flux reversal may be formed at different times after the initiation of reset and, again ϕ alone may not adequately account for the wall configuration. Thus the restriction to mmf which do not vary radically in time after the beginning of the reset process is necessary. Similarly, if ϕ' should change sign during reset, appreciable upswing of flux is likely to be accompanied by the formation of nuclei of positive flux reversal at the tape surfaces, and again ϕ would not be likely to provide adequate information about instantaneous core behavior. Thus the restriction to monotonic variation of ϕ with time is also necessary. Briefly, any attempt to express core behavior in terms of the variables F, ϕ', and ϕ should be restricted to rather "gentle" processes of flux reset from saturation with core voltages neither unusually large nor excessively fluctuating. These restrictions are fulfilled in many magnetic amplifier circuits; moreover, the use of variables F, ϕ', and ϕ to express core behavior appeared acceptable in reference 3, where data collected with core tests employing steps of mmf were used to predict mmf requirements under quite different conditions.

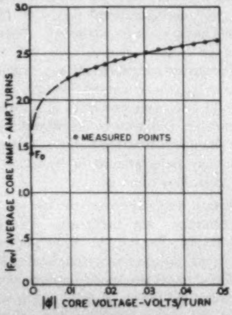


Fig. 2. F_{av} versus ϕ' test for core 1B

Experimentation for Functional Form and Parameters

The problem of expressing dynamic core behavior analytically is one of proper experimentation and interpretation to devise a core-reset function interrelating the instantaneous values of F, ϕ', and ϕ. In concept the simplest way of collecting information about the F, ϕ', ϕ relationship is to measure two of the variables throughout the range of interest for several fixed values of the third variable. This was done in reference 3 by resetting the core from saturation under steps of constant F and presenting the data graphically as ϕ' versus ϕ for various values of F; no analytical formulation of the data was made, however, mainly in view of the irregular shape of these curves.

Dynamic core properties can also be measured by resetting from saturation under constant ϕ' and measuring F versus t; since ϕ varies linearly with time at constant ϕ', these data are readily converted into F versus ϕ. This should yield essentially the same data as the constant-F test, but the constant ϕ' test, using square waves of core voltage, has certain advantages. (The constant mmf test holds no special relation to actual core operation; even in amplifiers where one winding carries a constant control current, the cores seldom experience a step of applied mmf, for there are usually coupled windings in which currents can flow, i.e., through "backfiring" rectifiers, thus modifying the value of total core mmf.) The more or less linear variation of mmf with time observed in the constant ϕ' test suggests rather extensive but not unreasonable simplifications for interpreting and accumulating core data. Thus, for a flux swing from $+\phi_s$, positive saturation, to $-\phi_s$ with ϕ and F linearly dependent upon time (and thus upon each other) the average resetting mmf can be shown to be equal to the mmf when flux is zero. Thus the average resetting mmf F_{av} can be measured under test with a d-c average-reading ammeter connected to carry only the reset current. The F, ϕ', ϕ relationship is then given by the sum

$$F = F_{av}(\phi') + a_1\phi$$

where a_1 could conceivably be a function of ϕ'.

It has been shown[4,5] that for the static properties of toroidal cores with high-grade magnetic materials the approximation that F varies linearly with ϕ is valid over a large part of the range between $+\phi_s$ and $-\phi_s$. In fact, if the static flux-mmf relationship is written as $F = a_0 + a_1\phi$,

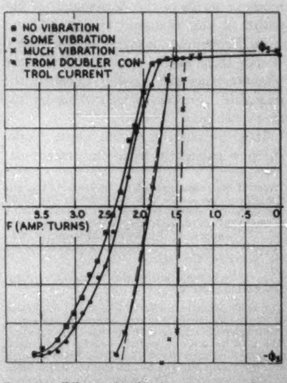


Fig. 4. Effect of vibration on static loop, core 2A

Table I. Core Properties

Core No.	Material; Manufacturer; Core Identification	Tape Thickness, Inches	ID,* Inches	OD,† Inches	Height, Inches	Turns	Wire Size	ϕsat, Webers	$-F_s$ AT‡	$a_1 = b_1$ AT/Webers	$-a_2$ AT/Volts, Turn	a_3	$-b_0$ AT	b_2 Turns/Ohm
1A	Orthonol Magnetics, Inc. 50020–2A	0.002	1.625	2.00	1.0	1,388 802 383 203	26	0.1478×10^{-3}	1.295	972	2.652	0.258	2.025	10.62
1B	Orthonol Magnetics, Inc. 50020–2A	0.002	1.625	2.00	1.0	1,388 802 383 203	26	0.1478×10^{-3}	1.410	1059	2.652	0.258	2.140	10.62
2A	Hypernik V Westinghouse Electric Corp.	0.002	1.5	2.5	0.5	954 569 386 208	26	0.1965×10^{-3}	1.906	2075	3.267	0.2064	3.13 3.2	10.40 8.89
2B	Hypernik V Westinghouse Electric Corp.	0.002	1.5	2.5	0.5	954 569 386 208	26	0.1965×10^{-3}	1.906	2075	3.481	0.2064	3.22	10.81
3A	Orthonol Magnetics, Inc. 50012–2A	0.002	1.250	1.75	1.0	1,100 1,100	27	0.1978×10^{-3}	1.32	1095	2.294	0.3043		
3B	Orthonol Magnetics, Inc. 50012–2A	0.002	1.250	1.75	1.0	1,100 1,100	27	0.1978×10^{-3}	1.32	1095	2.309	0.3096		
4	Orthonol Magnetics, Inc. 50012–2A	0.002	1.250	1.75	1.0	1,100 400	27	0.185×10^{-3}	0.852	767.6	1.815	0.234		

* Inside diameter. † Outside diameter. ‡ Ampere-turns.

a_1 is given in terms of the ratio x of inside to outside diameter of the magnetic tape build-up as

$$a_1=\frac{1-x}{1+x}\,\frac{a_0}{\phi_s} \tag{1}$$

Fluxmeter measurements on the several cores used in this work agree rather well with this theoretical value of a_1. No attempt is made in either of the references to derive a theoretical value of a_1 under dynamic conditions. However, data presented in reference 5 seem to indicate that the value of a_1 may increase under dynamic conditions up to about double its static value. However, for practical high-gain cores ($0.7<X<1.0$ approximately) the $a_1\phi$ term comprises a relatively small portion of the total mmf and itself arises from the approximation that the F, ϕ relationship is linear; hence it has been found convenient to assume a_1 equal to the value given by equation 1.

The tests to be described were all performed on cores with windings; thus effects of winding capacitance will be included in the data taken from dynamic tests. However, the coils were progressively wound and separated by tape and the number of turns was kept relatively low to keep capacitive effects small. Test data is not only easier to take when windings are provided but also appears to be more pertinent for actual prediction of the behavior of the wound core. On the other hand, if cores without windings are to be tested, a procedure can be devised for performing the prescribed test on an auxiliary core of known properties provided with convenient windings and coupled to the test core by means of a single turn coil or "permeameter cup"[6] of appropriately low resistance.

The F_{av} versus ϕ' test was performed using the circuit shown in Fig. 1; average resetting currents were measured for various amplitudes of the square wave of core voltage, meanwhile maintaining a constant flux swing slightly less than $2\phi_s$. The emf E in the reset branch is used, if necessary, to ensure that the core is driven into saturation; the dummy load is then used to avoid drawing a direct current component from the square waver. The variable-period square-wave voltage is obtained by shaping the output of a variable-frequency alternator; the shaping is performed by the saturable reactor circuit controlled from a constant current source yielding square-wave gate currents, the desired square wave of voltage appearing across R_0. The actual wave of core voltage with this circuit was "flat-topped" with voltage constant within 1.0% or 2.0% and with a rise time of about 1.0% or 2.0%

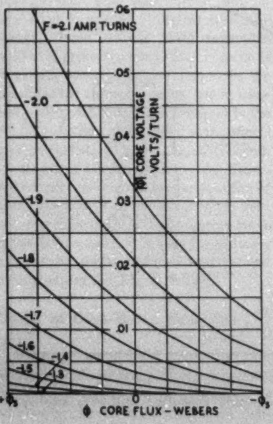


Fig. 5 (left). Plot of core reset function, ϕ' versus ϕ at constant F, core 3A

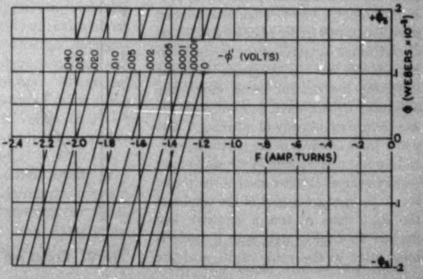


Fig. 6 (right). Reset branch of major hysteresis loops; plot of core reset function ϕ versus F at constant ϕ', core 3A

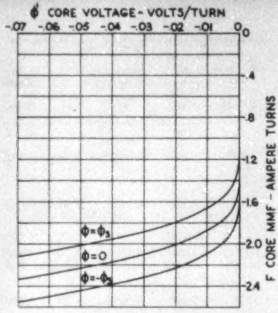


Fig. 7. Plot of core reset function F versus ϕ' for constant ϕ, core 3A

of the duration of the half-cycle. The test circuit operated satisfactorily over a frequency range of 15 to 100 cps. (An alternative means of generating the necessary voltage square waves is offered by transistorized magnetic inverters; here also a variable period and low internal impedance are not difficult to obtain.) The core voltage was measured with a rectifier-type average-reading a-c voltmeter. F_{av} was calculated from I_{av} readings with suitable corrections for any leakage of the rectifiers to obtain the half-cyclic average of resetting mmf.

A typical plot of the F_{av} versus ϕ' test data is presented in Fig. 2. Attempts to fit this data by power series were unsuccessful, but plotting the data in the form log $|F_{av}-F_0|$ versus log $|\phi'|$ led to good fit by a straight line (see Fig. 3) implying a functional form:

$$F_{av}=F_0+a_2(-\phi')^{a_3} \tag{2}$$

F_0 is the value of F_{av} when $\phi'=0$, i.e., the coercive mmf. For the resetting process ϕ', F_0, F_{av}, and a_2 are all negative. F_0 is not evaluated from the F_{av} versus ϕ' test. It could be found from static core tests with a fluxmeter or ballistic galvanometer, but due to the sensitivity of F_0 to core vibrations, Fig. 4, it can be evaluated more accurately under realistic conditions of core usage in terms of the control current of the doubler magnetic amplifier with very low control circuit impedance and using square-wave gate voltage. It was recognized already in reference 7 that under these conditions the postfiring control current corresponds to the static mmf requirement at the flux level of the preset core. In the log-log plot of the results of the F_{av} versus ϕ' test a straight line fits the data fairly well even if there is some error in the value of F_0, and the constants a_2 and a_3 adjust slightly, maintaining good fit of the function to the F versus ϕ' data except at very low values of $|\phi'|$. The complete core-reset function is then

$$F=F_0+a_1\phi+a_2(-\phi')^{a_3} \tag{3}$$

Values of the constants and data for the several cores tested are presented in Table I. Several ways of cross-plotting equation 3 are shown in Figs. 5, 6, and 7.

Simplified Reset Function

Use of the core-reset function (equation 3) in the analysis of amplifiers produces nonlinear differential equations because of the nonlinear relationship between F and ϕ'. Although such equations can be solved, in some cases analytically and in any case by numerical or graphical means, it may be desirable to formulate some simplified functional expression which will lead to more manageable circuit differential equations. In view of the large range of ϕ' in which the F_{av} versus ϕ' relationship departs only slightly from a straight line, Fig. 8, the linearization

$$F=b_0+b_1\phi+b_2\phi' \tag{4}$$

appears feasible when more accurate representation of core characteristics is not necessary, i.e., when only a small portion of the reset time of an amplifier is spent at low values of ϕ'. The variation of F with ϕ is not altered; hence $b_1=a_1$. However, b_0 arises purely from the approximation, not from any physical concept of coercive mmf as did F_0. The ampere turns b_0 and b_2 are determined from the zero intercept and slope respectively of the linearization as indicated in Fig. 8.

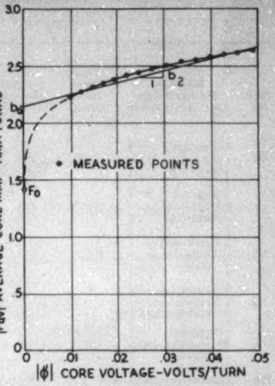


Fig. 8. Linear approximation of F_{av} versus ϕ' data for linearized core reset function

Processes Other Than Reset

The core-reset functions, being limited to reset processes, yield no information about the mmf requirements for ascending fluxes, for the saturated region, or for the transition from descending to ascending flux. Usually amplifiers are designed so as to minimize the effects of prefiring gate currents which flow while flux ascends toward saturation; thus these effects are

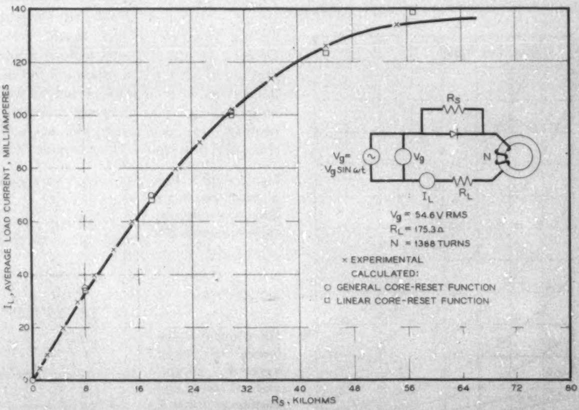


Fig. 9. Circuit and transfer characteristics of resistively controlled, half-wave amplifier

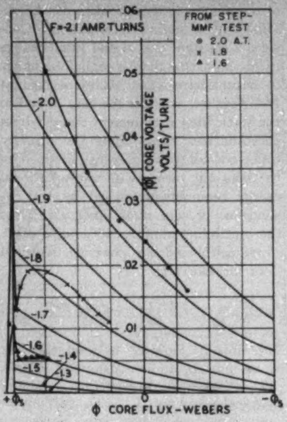


Fig. 10. Comparison of core reset function with results of step-magnetomotive force test, core 3A

often negligible or are satisfactorily approximated in some elementary way. On the other hand, the transition from descending to ascending flux, the bottom of the minor loop, although of limited interest in single-core amplifiers, may be quite important in the analysis of those 2-core amplifiers where a relatively large portion of the reset time is spent in this condition. Similarly, the top of the loop is relatively unimportant in single-core amplifiers at power frequencies. The processes which correspond to the variable slope in this region are thought to be more nearly reversible processes; hence static measurements may suffice when such information is needed. Furthermore, the energy stored during these reversible magnetic processes while gate current is increasing after firing is returned to the circuit as the current recedes, and in some cases where load volt-time integral is the circuit quantity of interest, these effects may go unnoticed. On the other hand, the storage of energy tends to delay the resetting process and some consideration of these effects may be necessary where very accurate knowledge of instantaneous wave forms is necessary or where the voltages induced by the changes of load current during saturation are appreciable.

Check of Core-Reset Functions

A quite significant check of the accuracy and adequacy of the core-reset functions in describing core behavior is afforded by the resistively controlled, half-wave amplifier shown in Fig. 9, for this is a case in which the proper account of ohmic drops during reset is most important. The analysis of the circuit will be deferred to Part II, but the results of the calculations of the transfer characteristics for both reset functions are presented in Fig. 9 showing excellent agreement between the calculated and experimental data.

Response to Step of MMF

The core-reset functions presented were obtained from tests performed under conditions of constant core voltage. It may also be of interest to check whether the reset functions obtained from such tests can be usefully applied to described processes of a somewhat different nature, namely, reset under a constant magnetomotive force F_k. The core-reset function then becomes

$$F_k = F_0 + a_1\phi + a_2(-\phi')^{a_3} \qquad (5)$$

This equation is plotted for several values of F_k and compared with experimental points from the corresponding step-mmf test in Fig. 10. Agreement is by no means perfect, but the discrepancy in values of ϕ' is less than 15% over most of the range of ϕ. The measured curves exhibit an initial spike of ϕ' due to the rapid change of mmf in the saturated region where the ϕ-versus-F slope is greater than zero due to some reversible magnetization of the core; in using the describing function this effect has been assumed away, so no such spike is calculated. The crest of ϕ' appears at ϕ_s when the core-reset function is used, whereas for the experimental data, the crest, aside from the initial spike, occurs at a slightly smaller value of ϕ. On the other hand, the postulation that reset under a step of mmf yields a core voltage which varies sinusoidally in time[8,9] implies that the curves of ϕ' versus ϕ of Fig. 11 would be ellipses with ϕ' being zero at $\phi = \pm\phi_s$ and a maximum at $\phi = 0$. Thus the curves of Fig. 10 are reassuring in that the agreement between calculations from reset functions based on constant-voltage tests and experimentation employing constant mmf is fair. Altogether, the postulated uniqueness of relationship between instantaneous values of F, ϕ', and ϕ appears reasonable for a rather wide range of processes.

Equation 5 can be solved readily for ϕ as a function of time during the constant mmf reset by solving explicitly for $d\phi/dt$, separating variables, and integrating. With the initial condition $\phi(t_0) = \phi_0$ this gives

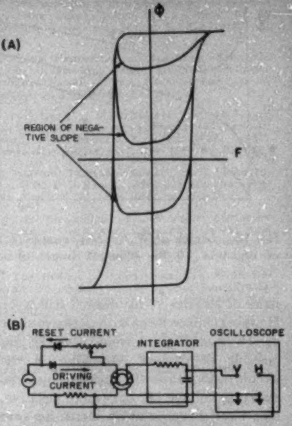


Fig. 11. A—Typical minor hysteresis loops with negative slope dϕ/dF of the resetting branch (Orthonol). B—Resetting circuit. (Reproduced by permission of Magnetics, Inc., from Catalog TWC100.)

$$\phi(t) = \frac{1}{a_1}\left\{F_k - F_0 + \left[(-F_k + F_0 + a_1\phi_0)^{\frac{a_3-1}{a_3}} + \frac{a_1(1-a_3)}{a_3(a_2)^{1/a_3}}(t-t_0)\right]^{\frac{a_3}{a_3-1}}\right\} \qquad (6)$$

and $\phi'(t)$ is found from equations 5 and 6:

$$\phi'(t) = \left\{\frac{a_1(1-a_3)}{a_2 a_3}(t-t_0) - [-\phi'(t_0)]^{a_3-1}\right\}^{\frac{1}{a_3-1}} \qquad (7)$$

Magnetic Hysteresis Loops

Common representations of the flux-mmf relationships of magnetic materials by hysteresis loops has led in some instances to the acceptance of these loops as the only possible locus of values of ϕ and F, or, in other instances, to the acceptance of certain major hysteresis loops taken under particular conditions of applied voltage or current as the basic core relationships. A plausible extension of this reasoning suggested that minor loops had the same sides as those of the major loop with top and bottom of the loop being

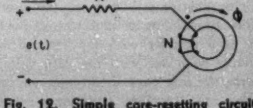


Fig. 12. Simple core-resetting circuit

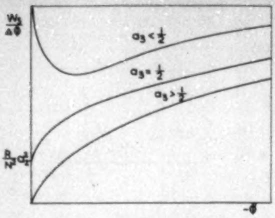


Fig. 13. Sketch of $W_s/(-\Delta\phi)$ versus $(-\phi')$ of equation 10 for different ranges of a_3

more or less flat at the desired flux level. However, minor loops with negative slope $d\phi/dF$ as in Fig. 11 would then be impossible to explain. Actually, in the circuit of Fig. 11 as the applied sinusoidal resetting voltage becomes more negative, core reset proceeds at an increasing rate requiring correspondingly more negative core mmf. As the resetting voltage levels off and becomes less negative the reset slows down and core mmf becomes less negative; as the voltage approaches zero the mmf passes through the static requirement and beyond to zero. Thus as flux decreases from $+\phi_s$ mmf has become first more, than less negative, giving rise to both positive and negative values of $\partial\phi/\partial F$ because of the changing values of ϕ'. The suggested core-reset functions offer approximations of basic core behavior which are consistent with this phenomenon.

With an appreciation of the effects of ϕ' upon dynamic mmf requirements it should not be suprising to find that the ϕ-F relationship for reset of cores in some practical circuits is quite irregular. The minor hysteresis loop with a "nick" at the bottom as presented in reference 7 for the doubler circuit is easily explained in terms of the rate of change of flux. Such irregularities of the reset branches of minor hysteresis loops are even more common in 4-core push-pull amplifiers where each core voltage may change greatly when one of the other cores saturates.[10] Similarly, wide fluctuations of core voltage and corresponding variations of resetting mmf can be expected in circuits which employ pulsing supplies.

For some particular purposes major dynamic loops obtained under special cyclic processes (e.g., sinusoidal voltages or sinusoidal mmf) may supply useful information. For instance, the core voltage of a doubler with low control impedance coincides with the gate voltage during the first part of the resetting half cycle, and therefore a dynamic major loop taken with a sinusoidal voltage of the same amplitude and frequency delineates the core mmf requirements during this part of the process.[11] But in more general cases the core voltage during reset may be of a quite different wave form, and such special loops, disguising the instantaneous dependence of F upon ϕ', are misleading.

Power and Energy Considerations

Toroidal tape-wound cores of modern high-permeability magnetic materials are inherently poor devices for storing energy in electromagnetic form, differently from ordinary linear chokes or reactors. As core flux is cyclicly varied by alternately positive and negative applied voltages, the instantaneous power input to the reactor does not change sign (energy always flows into the core) except for the small changes of flux associated with reversible magnetization in the saturated region or in crossing the static loop from one branch to the other. The amounts of energy involved in these reversible effects are usually quite small with respect to the core losses and with respect to the I^2R losses of the associated circuit, and therefore will be considered negligible for this discussion. The hysteresis effects in an electromagnetic system have as mechanical analog Coulomb friction, which also consistently opposes and absorbs energy from the motivating force, never returning energy to the system except as heat. Moreover, the conversion of electromagnetic energy to heat energy occurs on an instantaneous basis as the irreversible eddy current and spin relaxation effects which correspond to a given ϕ and ϕ' take place.

In magnetic amplifier circuits such energy considerations may be important both from the aspect of core heating and of energy requirements of control. For the simple series circuit of Fig. 12 containing a reactor with N turn winding, an emf source of $e(t)$ volts, and a resistor R ohms instantaneous input power is

$$\begin{aligned} p(t)&=e(t)i(t)=i[Ri+N\phi'] \\ &=\frac{R}{N^2}F(t)^2+\phi' F(t)=P_{\text{resistor}}+P_{\text{core}} \end{aligned} \quad (8)$$

where $F(t)=Ni(t)$. Solutions of the circuit for $\phi(t)$ and $\phi'(t)$ with the aid of the core-reset function can be used for power and energy calculations.

The question arises whether in resetting a core by specified change of flux $\Delta\phi$, there may not be optimum values of ϕ' and the corresponding reset time to minimize either the energy dissipated by the resistor or by the core, or the energy output of the emf source. Some indication of the answer to this problem can be obtained in a simplified situation. Consider a constant emf E and small charge in flux $\Delta\phi$ so that there is negligible change in the mmf term $a_1\phi$ and therefore constant ϕ'. Further, consider the core biased to the static mmf requirement $F_0+a_1\phi$ by a constant bias mmf. Here the primary concern is not with the energy supplied by the bias but rather with the additional energy requirement caused by the dependence of core mmf upon ϕ'. Then the controlling mmf F_c is $F_c=a_2(-\phi')^{a_3}$, where $a_2<0$, and the power p_s required from the source is

$$P_s=\frac{R}{N^2}\,a_2^2(-\phi')^{2a_3}+(-a_2)(-\phi')^{a_3+1} \quad (9)$$

The time Δt required to reset the flux by an amount $\Delta\phi$ at constant ϕ' is $\Delta t=(\Delta\phi)/\phi'$, and the energy per unit flux change required for the reset is $W_s/(-\Delta\phi)$:

$$\begin{aligned} \frac{W_s}{(-\Delta\phi)}&=P_s\frac{\Delta t}{(-\Delta\phi)}=(-a_2)\left[\frac{R}{N^2}(-a_2)\times (-\phi')^{2a_3-1}+(-\phi')^{a_3}\right] \quad (10) \\ &=(-a_2)(-\phi')^{a_3}\times \left[\frac{R}{N^2}(-a_2)(-\phi')^{a_3-1}+1\right] \end{aligned}$$

The energy dissipated in the core $(-a_2)(-\phi')^{a_3}(-\Delta\phi)$, increases as $|\phi'|$ increases. The energy absorbed by the resistor, $R/N^2\, a_2^2(-\phi')^{2a_3-1}(-\Delta\phi)$, increases with $|\phi'|$ if $a_3>1/2$ but decreases if $a_3<1/2$. The functional relationship of $W_s/(-\phi)$ to $(-\phi')$ is sketched in Fig. 13 for a_3 less than, equal to, and greater than 1/2, $(a_3>0)$. For $a_3\geq 1/2$ the energy required from the source decreases as $|\phi'|$ decreases, but for $a_3<1/2$ (which is the range found for the seven cores tested) there is an optimum ϕ',

$$(-\phi')_{\text{opt}}=\left[\frac{R}{N^2}(-a_2)\left(\frac{1}{a_3}-2\right)\right]^{\frac{1}{1-a_3}} \quad (11)$$

From the plot of equation 10 for core 1A, Fig. 14, it is seen that the minimum of control source energy extends over a rather wide range of ϕ' for the larger values of R/N^2, and that little additional energy is required for core voltages greater in absolute value than the optimum. At lower resistances the ratio of source energy at large $|\phi'|$ to the optimum energy varies from 1 up to about 5. At extremely low values of $|\phi'|$ the energy requirement is much greater, making reset at such low rates uneconomical of energy. Moreover, in any practical circuit, bias to the exact static requirement for any considerable range of values of resetting flux is impossible, and such additional mmfs as are necessary to satisfy changes in static re-

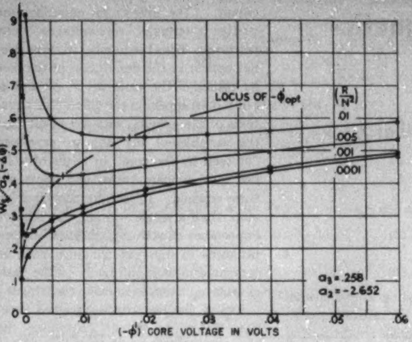


Fig. 14. Source-energy requirement versus rate of core reset for various R/N^2 from equation 10 for core 1A

quirement with flux level will cause increased i^2R losses. These losses become increasingly important as $|\phi'|$ decreases because of the correspondingly longer reset time during which they occur. Thus it seems that where there is a possible choice for wave form of resetting voltages, the choice should be directed away from very small voltages and long resetting times. But in instances where core heating is a problem, some advantage may be found in resetting at fairly low rates of change of core flux to reduce core losses; whether this will result in any appreciable reduction of temperature rise of the reactor will certainly depend also upon other practical considerations such as how much of the resistive power loss occurs in the windings and what is the lower limit on practical values of $|\phi'|$.

Summary

An empircial relationship between instantaneous values of magnetic variables pertinent to the reset of grain-oriented nickel-iron tape cores from initial saturation has been presented in the form of core-reset functions. Care has been taken to keep this information consistent with the physical understanding of the dynamic processes of flux reversal in the magnetic tape and to point out the limitations thus imposed upon its applicability. The postulation that the variables F, ϕ', and ϕ are adequate to express the core reset under certain restrictions has been checked and confirmed. It is recognized only too well that the core-reset functions as developed here do not arise from any general physical theories, and that the constants are not evaluated in terms of basic physical constants; moreover, some error is introduced by the methods and approximations which were found expedient in developing the functions empirically. Yet the formalization of the description of dynamic core behavior allows and motivates more precise thinking and expression of pertinent ideas, adds to the general understanding of the behavior of magnetic amplifier cores, and gives a more concrete basis for magnetic-amplifier analyses.

List of Symbols

a_i—parameters of general core-reset function
a_0—ampere-turns
a_1—ampere-turns per weber
a_2—ampere-turns per volt^{a_3}
a_3—dimensionless
b_i—parameters of linear core-reset function
b_0—ampere-turns
b_1—ampere-turns per weber
b_2—ampere-turns per volt
F—instantaneous mmf, ampere-turns
F_{av}—half-cycle average mmf for reset, ampere turns
F_k—amplitude of step of mmf, ampere-turns
F_0—coercive mmf, ampere-turns
N—number of turns of winding
p—instantaneous power, watts
R—resistance, ohms
t—time, seconds
x—ratio of inside diameter to outside diameter of magnetic tape build-up
W—energy, watt-seconds
ϕ—instantaneous flux, webers
ϕ'—instantaneous time rate of change of core flux, volts = webers per second
ϕ_s—saturation flux, webers

References

1. Flux Reversal in Magnetic Amplifier Cores, F. J. Friedlaender. *AIEE Transactions*, vol. 75, pt. I, July 1956, pp. 268–78.

2. Flux Resetting Characteristics of Several Magnetic Materials, Hoobert Huhta. *Ibid.*, vol. 73, pt. I, May 1954, pp. 111–14.

3. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *Ibid.*, vol. 76, pt. I, May 1957, pp. 229–40.

4. Static Properties of Toroidal Magnetic Cores, H. M. McConnell. *Magnetic Amplifiers Technical Report No. 13*, Carnegie Institute of Technology, Pittsburgh, Pa., 1953.

5. Influence of ID/OD Ratio on the Static and Dynamic Magnetic Properties of Toroidal Cores, R. W. Roberts, R. I. Van Nice. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 599–607.

6. Testing Toroidal Cores, L. J. Johnson, G. J. Maki, S. E. Rauch. *Electrical Manufacturing*, New York, N. Y., May 1954, pp. 130–35.

7. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, Harold W. Lord. *AIEE Transactions*, vol. 72, pt. I, Jan. 1954, pp. 721–28.

8. The Effect of Eddy Current on the Behavior of Magnetic Amplifiers, John Travis. *Ph.D. Thesis*, Purdue University, Lafayette, Ind., June 1955.

9. Magnetic Amplifier Analysis Using a Generalized Model for the Saturable Reactor Core, Herbert H. Woodson. *WADC Technical Note 56-198*, Massachusetts Institute of Technology, Cambridge, Mass., May 1956.

10. An Analysis of Magnetic Amplifiers in Push-Pull, D. L. Critchlow. "Conference on Magnetic Amplifiers in Push-Pull," *AIEE Special Publication T-86*, July 1956, pp. 52–73.

11. The Operation of Magnetic Amplifiers with Various Types of Loads, Part II, L. A. Finzi, R. R. Jackson. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 279–88.

◆

Discussion

H. W. Lord (General Electric Company, Schenectady, N. Y.): This paper defines the symbol F_0 as coercive mmf ampere-turns. However, the value used for F_0 in the analysis is not determined from a major d-c hysteresis loop but from a value of current flowing in the control winding during the postfiring interval in a certain type of magnetic-amplifier circuit.

Since the widely accepted meaning of coercive force H_c is quite different from that actually used for F_0 in this paper, I feel that the paper should define F_0 so as to distinguish it from H_c, the coercive mmf as determined by a major hysteresis loop.

G. C. Feth: Following is a more precise definition of F_0: F_0 for a magnetic core is the value of mmf required by the core in the limit as ϕ' goes to zero at $\phi = 0$ under the condition of reset from saturation in a cyclic magnetization process performed at the particular frequency of interest (here 60 cycles per second).

For cores not sensitive to vibration there may be no noticeable difference between the values of F_0 and d-c coercive mmf. This was actually the case for some of the cores tested. However, a core with one of the greatest per-cent differences between F_0 and d-c coercive mmf was chosen for the data of Fig. 4. For such a core the value of F_0 is quite different from statically measured values and it is wise to make a clear distinction between F_0 and d-c coercive mmf.

Core Reset Functions in Magnetic-Amplifier Analysis

Part II—Determination of Amplifier Performance

G. C. FETH
ASSOCIATE MEMBER AIEE

A CRITICAL analysis of most practical magnetic amplifiers requires a detailed knowledge of the behavior of the magnetic core. For many amplifiers the analysis may be conveniently separated into two parts: determination of the load voltage or current in terms of the extent of the previous reset of core flux from saturation, and determination of the flux reset in terms of the control or input signal. The output may often be evaluated in terms of the preset flux level with only a meager knowledge of core properties, but to determine the value of the controlling variable needed for a given flux reset or to ascertain the effects and optimum values of circuit parameters requires a more extensive understanding of core behavior. The previously published Part I[1] of this paper supplied core-reset functions approximating the relationships between the variables, core flux ϕ, mmf (magnetomotive force) F, and rate of change of flux ϕ', for the process of reset from saturation under certain conditions. In this paper the core-reset functions are contrasted with some other core approximations, several examples of the use of core-reset functions to calculate amplifier performance are shown, and an example of a model theory for the design of analog computers simulating magnetic-amplifier systems is presented.

Reset for Various Core Approximations

Numerous simplifying approximations of magnetic-core behavior are reported in the literature. In Table I five core approximations with various degrees of complexity are presented, and for the simple case of reset from a voltage source in series with a resistor, the differential equations for ϕ as a function of time are formulated and solved, assuming the core to be unsaturated. The applicability of these approximations will be discussed and some particular implications will be examined.

Many magnetic amplifiers are analyzed conveniently in terms of controlling voltages, acting in the control circuit during the reset half-cycle. When the currents which flow to fulfill the mmf requirements of the cores do not produce appreciable voltage drops, no need may be felt for expressing these requirements with any great degree of accuracy; in such cases even the 3-zone idealization, Table I(a), may be satisfactory. However, a better approximation of the core properties is necessary, for instance, to obtain even an estimate of the required current-carrying capacity of the control windings or the power requirements of the control source. For such purposes a slightly less oversimplified idealization such as the square-loop with fixed width, as in Table I(b), may be adequate. With this approximation there arises the practical question of what value to assign to b_0, the half-width of the loop, in order to reasonably approximate the core behavior, for the width of such a loop may well vary by a factor of two or more, depending upon the conditions under which it is measured. Furthermore, both the square-loop and the 3-zone idealizations are inadequate to describe the reset of a core by means of a current source, and any attempt to use them in such analyses is likely to yield highly misleading results. Thus, for general problems of core reset where appreciable voltage drops result from the currents which flow to fulfill the dynamic mmf requirements of the core, more realistic approximations accounting for the effects of ϕ' are needed.

Core approximations which introduce the effects of ϕ' are presented in Table I(c), (d), and (e). Of course, as the approximations become more realistically descriptive of core behavior, the parameters of the core and circuit are more in evidence in the analytical expressions; for the 3-zone idealization N was the only parameter which appeared, whereas for the general core-reset function there are four parameters of the core and two of the circuit. (The differential equations are written in terms of "core voltages," volts per turn. Thus the number of turns does not appear in the description of the core by means of the variables F, ϕ', and ϕ, but it does appear in the differential equations of the circuits, with the result that the variables of any loop are expressed "as seen by the core" in terms of volts per turn, ampere turns, and ohms per turns squared. In fact, for cores bearing more than a single winding the relative importance of voltages, currents, and impedances in different windings may be compared meaningfully only on the basis of such reduction to common turns basis.)

Use of the linearized core-reset function, Table I(c) or (d), containing the parameter b_2, which may be evaluated by direct measurements on the core, establishes a criterion as to the relative magnitude of the impedance of the resetting circuit, and makes possible more meaningful attempts to match the impedance of the signal source plus control winding resistance to an "apparent input impedance" of the core. This is seen directly from equation 6 of Table I; here, in the term (b_2+N^2/R), the importance of the R/N^2 term is determined relative to a fixed property of the particular core, namely $1/b_2$. (The only other term in which R/N^2 appears is in the numerator $[(E/N \div R/N^2)-b_0]$; R/N^2 is associated with E/N since both quantities are parameters of the circuit whereas b_0 is a parameter of the core. Thus, to compare the behavior of the core at different values of R/N^2, the ratio $E/N \div R/N^2$ must be kept constant. This point of view appears reasonable also since $E/N \div R/N^2$ is a significant physical quantity, the mmf acting on the core when there is no flux change.) This relationship of control circuit resistance to $1/b_2$ might, perhaps, also be a useful parameter in some amplifier circuits for determining the relative suppression or constraint of voltages induced in the control loop. Unfortunately, no such simple

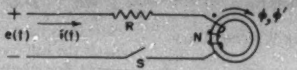
---

Paper 57-921, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957, and re-presented at the Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 13, 1957; made available for printing December 9, 1957.

G. C. FETH is with the General Electric Company, Schenectady, N. Y.

This work has been performed in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology. The work has been supported both by the General Electric Educational and Charitable Fund through a Coffin Fellowship and by the United States Office of Naval Research. The author wishes to acknowledge the valuable guidance of his thesis supervisor, Prof. L. A. Finzi of Carnegie Institute of Technology.

Table I. Voltage Reset for Various Core Approximations

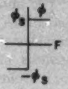


$$E_{av} = \frac{1}{T}\int_0^T e\,dt \qquad \text{S CLOSES AT } t=0\text{; OPENS AT } t=T$$

a. THREE–ZONE IDEALIZATION: $F=0$ FOR $-\phi_s \le \phi \le \phi_s$

$$\phi' = \frac{e}{N} \tag{1}$$

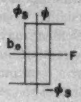
$$\Delta\phi = T\,\frac{E_{av}}{N} \tag{2}$$



b. SQUARE LOOP WITH FIXED WIDTH: $F=b_0$ FOR $-\phi_s \le \phi \le \phi_s$

$$\phi' = \frac{R}{N^2}\left[\frac{e/N}{R/N^2} - b_0\right] \tag{3}$$

$$\Delta\phi = T\,\frac{R}{N^2}\left[\frac{E_{av}/N}{R/N^2} - b_0\right] \tag{4}$$

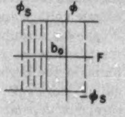


c. LINEAR CORE–RESET FUNCTION WITH $b_1\phi = 0$:

$$F = b_0 + b_2\phi' \text{ FOR } -\phi_s \le \phi \le \phi_s \tag{5}$$

$$\phi' = \frac{\frac{R}{N^2}\left[\frac{e/N}{R/N^2} - b_0\right]}{1 + b_2\,R/N^2} = \frac{\left[\frac{e/N}{R/N^2} - b_0\right]}{b_2 + N^2/R} \tag{6}$$

$$\Delta\phi = \frac{T\,\frac{R}{N^2}\left[\frac{E_{av}/N}{R/N^2} - b_0\right]}{1 + b_2\,R/N^2} \tag{7}$$



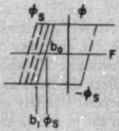
d. LINEAR CORE–RESET FUNCTION:

$$F = b_0 + b_1\phi + b_2\phi' \text{ FOR } -\phi_s \le \phi \le \phi_s \tag{8}$$

At $e = E =$ CONSTANT

$$\phi' + \frac{b_1}{b_2 + N^2/R}\,\phi = \frac{\frac{E/N}{R/N^2} - b_0}{b_2 + N^2/R} \tag{9}$$

$$\Delta\phi = \frac{1}{b_1}\left[\frac{E/N}{R/N^2} - b_0 + b_1\phi_s\right]\left[1 - \exp\left(\frac{-b_1 T}{b_2 + N^2/R}\right)\right] \tag{10}$$



e. GENERAL CORE–RESET FUNCTION:

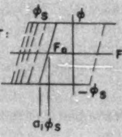
$$F = F_0 + a_1\phi + a_2(-\phi')^{a_3} \text{ FOR } -\phi_s \le \phi \le \phi_s \tag{11}$$

$$\phi' + (R/N^2)a_2(-\phi')^{a_3} + (R/N^2)\,a_1\phi = E/N - (R/N^2)F_0 \tag{12}$$

At $t=0$, $\phi(0) = \phi_s$, $\phi'(0) = \phi'_0$; THEN FOR $e = E =$ CONSTANT:

$$t = \frac{1}{a_1\frac{R}{N^2}}\ln\frac{\phi'_0}{\phi'} - \frac{a_2\,a_3}{a_1(1-a_3)}\left[(-\phi')^{a_3-1} - (-\phi'_0)^{a_3-1}\right] \tag{13}$$

$$\phi = \frac{1}{a_1}\left[\frac{E/N}{R/N^2} - F_0 - a_2(-\phi')^{a_3} - \frac{\phi'}{R/N^2}\right] \tag{14}$$



relationship results from the use of the general core-reset function since no explicit expression for flux reset is forthcoming.

When the value of F depends upon ϕ in the unsaturated core, Table I(d) and (e), the amount of reset depends upon the instantaneous values of applied emf (electromotive force); when F is assumed independent of ϕ the amount of reset depends only on the average emf. For high-gain cores with large inside diameter to outside diameter ratios, the dependence of F on ϕ is not very large, yet its effects are noticeable in some instances. Table I(c) and (d) compare the two effects of neglecting or retaining the variation of F with ϕ. It is, in large part, this dependence of the amount of reset upon the waveforms of applied voltage which causes any application of core information as obtained from a control characteristic under one set of circuit conditions to be of doubtful value for any general analytical purposes. The equations of Table I(d) with b_2 set equal to zero apply for a hysteresis loop with parallel sides of less than infinite slope; however, a 3-zone approximation with less than infinite permeability is not considered here, for it bears no more useful relation to actual characteristics of tape-wound cores than the other approximations and it may actually be quite misleading.

The nonlinear differential equation 12 of Table I(e) describes the reset of a core approximated by the general core-reset function. For a constant applied emf the solution of this equation is obtained by differentiating once with respect to time, separating variables, and integrating. The results are presented in equations 13 and 14; these equations cannot be solved explicitly for ϕ and ϕ' because of the transcendental functional form, but the results may be plotted readily from these two equations for any given values of the circuit parameters. Any general plot or nomogram is excluded because of the large number (7) of independent parameters: F_0, a_1, a_2, a_3, R/N^2, E/N, and ϕ_s.

Although the general core-reset function does not readily lead to simple analytical expressions from which general conclusions may be drawn, yet its use is quite worth while in terms of the accuracy of representing core behavior, especially where large ranges of ϕ' are involved.

As the resistance of the resetting circuit increases, the effects of ϕ' on the mmf requirement become increasingly important. For reset under constant mmf F_K the solutions for $\phi(t)$ and $\phi'(t)$ for the general core-reset function are presented in reference 1; for the linear function the results are obtained from equations 9 and 10 of Table I by replacing $E/N \div R/N^2$ by F_K and N^2/R by zero.

Use of Core-Reset Functions in Magnetic-Amplifier Analysis

In order to analyze a physical system one undertakes to apply physical laws to the system, to express the behavior of each element in terms of suitable approximations, and to determine the mathematical relationships which obtain between the significant variables. In a magnetic amplifier both the magnetic cores and the rectifiers are approximated by discontinuous relationships between the variables, i.e., a rectifier, described in terms of the relationship between current and voltage, has different parameters in the conducting and nonconducting directions, and a core, whose saturation is determined by the level of magnetic flux, is approximated in different ways depending upon whether it is saturated or unsaturated, upon the sign of ϕ', and upon its electromagnetic history. Thus, for analysis, the operation of a magnetic amplifier is separated into intervals during which continuous relationships exist between all variables; the operation is described mathematically for each of these intervals, and the conditions for termination of each interval are established. This completes the mathematical description of the system; from this the steady-state behavior may be found or, with given initial conditions, transient problems may be solved. Such analysis is not particularly simple: numerical analysis may be resorted to,

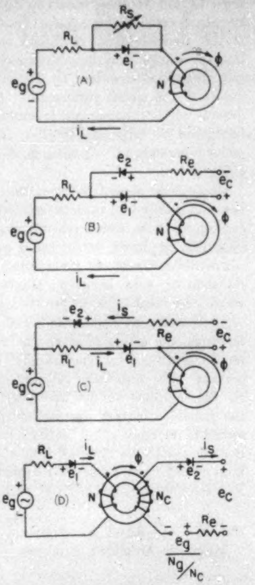


Fig. 1. Single-core single-winding half-wave amplifiers

and the actual solution may well be performed with the aid of digital computers. In some cases, a fair approximation of the features of interest in a particular system may be made without such a complex description of the system.

Here, the approach will be to give some examples of the use of core-reset functions in analysis and to demonstrate the accuracy with which amplifier performance may be calculated. This work is not intended as a detailed analytical treatment of any particular amplifier circuits. Rather, several common circuits will be considered, their various intervals or modes of operation enumerated, the formulation and solution of the mathematical problem described briefly, and the results presented and discussed. Where worth-while simplified approximations of the over-all behavior are forthcoming, they will be described and the results presented. Only resistive loads will be considered, for the problem of determining the effects of the load upon the output is often a fairly straightforward problem[2] which may be treated separately from the problem of determining the amount of flux reset in terms of the control quantity, the latter being of primary interest here since the core behavior is involved directly.

SINGLE-CORE SINGLE-WINDING HALF-WAVE AMPLIFIER

The single-core single-winding amplifier shown in Fig. 1(A), (B), and (C) provides a convenient starting point for analysis of magnetic amplifier circuits, for it avoids problems of coupling between windings and the attendant possibility of recifier "backfiring." The controlling element may be either a signal voltage or a variable impedance, e.g., a photoresistor, or a combination of both. In the operation of the amplifier the output volt-time integral during a gating half-cycle is the difference between the half-cyclic integral of gate voltage and the volt-time integral $N(\phi_s-\phi_p)$ absorbed by the core during reset, where ϕ_p is the preset flux level found at the beginning of that particular gating half-cycle. The rate of reset of core flux depends instantaneously upon the other voltages appearing in the circuit: the supply voltage, the signal voltage, voltages appearing across rectifiers, resistors, or other components. In some particular instances the voltages that affect the reset most significantly do not depend strongly on the currents flowing in the loops; in such cases, of course, no accurate knowledge of the values of these currents is required. But in the situation of Fig. 1(A) control depends basically upon the voltage drop of the resetting current in the controlling resistor R_s, and the voltage equation of the resetting loop is

$$N\phi'=e_g-(R_L+R_s)i \qquad (15)$$

Expressing the instantaneous core mmf requirement by means of the general core-reset function, equation 11, Table I, leads to the following differential equation of reset for sinusoidal gate voltage, $e_g=E_g \sin \omega t$:

$$\phi'+a_2\frac{R_L+R_S}{N^2}(-\phi')^{a_3}+a_1\frac{R_L+R_S}{N^2}\phi =\frac{E_g}{N}\sin \omega t-\frac{R_L+R_S}{N^2}F_0 \qquad (16)$$

This is a first-order, nonlinear differential equation, and may be solved for given values of the parameters by numerical or graphical means. (In particular, a step-by-step numerical solution for ϕ can be found by assuming successive values of ϕ', calculating for each value of ϕ' the corresponding value of ωt with the use of the original flux value, and recalculating ωt with an intermediate value of flux.) Thus ϕ_p can be calculated, and the average value of load current is

$$I_L=\frac{E_g}{\pi R_L}-\frac{\omega N(\phi_s-\phi_p)}{2\pi R_L}-\frac{E_f}{2R_L}-\frac{|I_{\text{reset}}|}{2} \qquad (17)$$

where E_f is the rectifier forward drop and I_{reset} is the half-cyclic average of resetting current flowing through the load in the reset half-cycle. These corrections are often quite small. In Fig. 2 the comparison of the results of these calculations with the measured transfer characteristic of a given amplifier shows excellent agreement.

Use of the linear core-reset function, equation 8 of Table I, yields the following linear differential equation of reset

$$\phi'+\frac{b_1}{b_2+\dfrac{N^2}{R_L+R_S}}\phi=\frac{E_g/N}{1+b_2\dfrac{R_L+R_S}{N^2}}\times \sin \omega t-\frac{b_0}{b_2+\dfrac{N^2}{R_L+R_S}} \qquad (18)$$

which may be solved, with the proper initial conditions, for the preset value of flux ϕ_p existing at the termination of reset near the end of the reset half-cycle. I_L is then evaluated from equation 17. The calculated points of the transfer characteristic with this core approximation also show quite good agreement with the experimental data; see Fig. 2.

If a square hysteresis loop with fixed width, Table I(B), is assumed for the core, the differential equation of reset becomes

$$\phi'=\frac{E_g}{N}\sin \omega t-h_0\frac{R_L+R_S}{N^2} \qquad (19)$$

This equation may be integrated directly and ϕ_p evaluated. There is some uncertainty as to the value of h_0 which is to be used with the square-loop approximation. Calculations were made for three possible values of h_0, i.e., h_0 corresponding to the coercive mmf F_0, to the width of the 60-cycle normal-sinusoidal-voltage dynamic loop at zero flux; and for $h_0=b_0$. The variation in predicted output under use of $h_0=F_0$ or h_0 equal to the dynamic 60-mmf requirements was in the ratio of about 1 to 2, with the actual experimental value lying between these two limits, as would be expected. In this instance it was found that the values calculated for $h_0=b_0$ give good agreement over a large part of the range of control. This may well be partly the result of cancellation between errors caused by neglecting the variation of F with ϕ and of F with ϕ', which errors produce opposite effects over part of the range. If the square-loop approximation is to be used, it seems that the value chosen for h_0 might just as well be selected as b_0, i.e., the value of F found by extra-

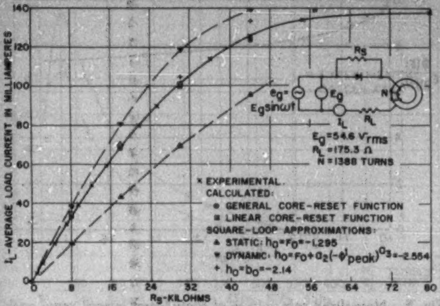


Fig. 2. Transfer characteristic of resistively controlled single-core single-winding half-wave amplifier for normal voltage, core 1A

polating the F_{avg} versus ϕ' curve to zero ϕ' by means of a straight line approximation of the curve at large ϕ'; see reference 1. Evidently the choice of the type of core approximation to be used depends on the accuracy of predictions needed in any particular instance. It may well be recognized that any idealization of core behavior different from the trivial 3-zone representation involves certain analytical complications.

Some single-core half-wave amplifiers employing a separate winding for control have the same differential equations as a single-winding amplifier (if perfect coupling between windings is assumed). For instance, in the amplifiers of Fig. 1(C) and (D), by comparing corresponding equations of the gate loop, control loop, and mmf in the two cases it can be shown that, for identical gate circuit parameters the two circuits are identical if

$$e_c^* = (N_c/N)e_c$$

$$e_s^* = (N_c/N)e_s$$

$$i_s^* = (N/N_c)i_s$$

$$R_s^* = (N_c^2/N^2)R_s$$

[where the variables with * refer to Fig. 1(D) and the others to Fig. 1(C)], these being the reflected-value transformations of ordinary transformer techniques. Thus the same analysis as for the single-winding amplifier applies also to such circuits. The many similar circuits, built up with more windings and perhaps more cores, where voltages are added and subtracted in various ways to perform the desired control, may also be analyzed with the use of the core-reset function to supply the necessary information about the currents and associated *IR* drops resulting from the core mmf requirements. In particular, analyses may be made for the practical case of source impedances different from zero.

SINGLE-CORE MULTIPLE-WINDING HALF-WAVE AMPLIFIER

As an example of the single-core multiple-winding half-wave amplifier the operation of the 2-winding amplifier of Fig. 3 with constant control voltage E_c and control resistance R_c will be considered. In the positive half-cycle of gate voltage the core is brought to saturation, firing at the angle α. The core remains saturated until the angle β $(\beta \le \pi)$ when the core mmf, $N_gi_g+N_ci_c$, (with i_c negative) becomes equal to the static mmf requirement of the resetting branch. The rectifier is still in the state of forward conduction; thus the voltages appearing in both control and gate loops affect the core reset. At the angle $\gamma(\gamma \ge \pi)$ the rectifier voltage becomes negative and the rectifier blocks. In the interval $\pi \le \omega t \le \gamma$ the gate current is quite small; prior to π the core voltage is practically zero, and after π it is limited essentially to e_g/N_g(i.e., one may say that the core voltage is "gate-limited" and the interval $\beta \le \omega t \le \gamma$ is the first interval of gate-limited reset). After γ the core is reset under the influence of the control circuit alone until the rectifier voltage again becomes positive at δ $(\delta \le 2\pi)$; this marks the beginning of a second interval of gate-limited reset. The core continues to reset with core voltage practically equal to e_g/N_g (assuming that negative saturation is not reached) until the gate voltage again begins raising the flux toward saturation; this occurs at some time after 2π. In steady state the cyclic-average load current I_L may be found by integrating the gate circuit differential equation to obtain

$$I_L = \frac{E_g}{2\pi R_L}(\cos\delta - \cos\gamma) + \frac{\omega N_g}{2\pi R_L} \times [\phi(\delta) - \phi(\gamma)] - \frac{E_f}{2\pi R_L}(\pi + \gamma - \delta) \quad (20)$$

for $E_g \le E_{gN} = \omega N_g\phi_s$. Between γ and δ core voltage may be determined from the control-loop voltage equation and the core-reset function; the initial conditions are: $\phi'(\gamma) = (E_g/N_g)\sin\gamma$, $\phi(\gamma) = \phi_s - (E_g/\omega N_g) \times (1+\cos\gamma)$. For large values of R_c/N_c^2 and small amounts of reset $|\phi'|$ will be small, and to avoid introducing errors the general core-reset function should be used. Thus equation 12 of Table I is the differential equation for reset by control, $\gamma \le \omega t \le \delta$. The angle γ is evaluated by substituting the values of $\phi'(\gamma)$ and $\phi(\gamma)$ into this equation and solving the resulting transcendental equation for γ:

$$-\sin\gamma = \frac{R_c}{N_c^2}\left[a_3\left(\frac{E_g}{N_g}\right)^{a_3-1}(-\sin\gamma)^{a_3} + \frac{a_1}{\omega}\left(\frac{\omega N_g\phi_s}{E_g} - (1+\cos\gamma)\right) + \frac{N_gF_0}{E_g}\right] - \frac{E_c}{N_c}\,\frac{N_g}{E_g} \quad (21)$$

Solution of the differential equation and use of the defining relationship

$$\phi'(\delta) = \frac{E_g}{N_g}\sin\delta \quad (22)$$

yields the following transcendental equation for δ:

$$\delta = \gamma + \frac{\omega}{a_1}\,\frac{a_3(-a_2)}{1-a_3}\left(\frac{E_g}{N_g}\right)^{a_3-1}[(-\sin\delta)^{a_3-1} - (-\sin\gamma)^{a_3-1}] + \frac{N_c^2}{R_c}\,\frac{\omega}{a_1}\ln\left(\frac{\sin\gamma}{\sin\delta}\right) \quad (23)$$

$\phi(\delta)$ can be evaluated from equation 12 and I_L from equation 20. A few calculations on the basis of these equations show quite close agreement with the experimental control characteristics shown in Fig. 4.

In view of the fairly good linearity of the control characteristics presented in Fig. 4, it appears worth while to concentrate some effort on estimating two points which will determine a satisfactory straight-line approximation. The control mmf which just produce full output and quiescent output might at first glance appear to be the simplest to approximate. However, at the verge of resetting, $F = F_0 + a_1\phi_s$, and this point as plotted in Fig. 4 at full output does not lie satisfactorily close to the linear approximations of most of the characteristics. Poor agreement

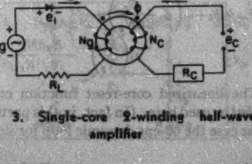


Fig. 3. Single-core 2-winding half-wave amplifier

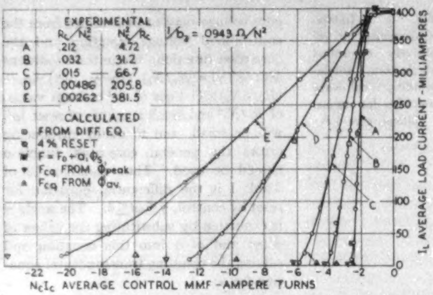


Fig. 4. Control characteristic of single-core 2-winding half-wave amplifier with various R_c/N_c^2 and normal voltage, core 1A

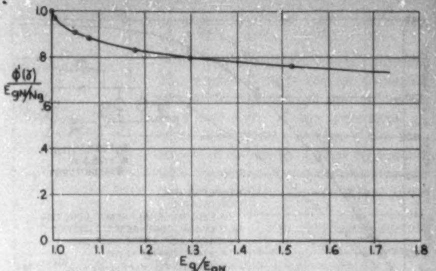


Fig. 5. Normalized core voltage of control-reset interval versus peak gate voltage for $E_g \geq E_{gN}$

here is not surprising, for this is in the range of very small reset from saturation where the core-reset function is not a very accurate approximation (since any curvature of the top of the loop has been ignored). This difficulty may be bypassed by making a rough approximation of the mmf required to reset the flux by a few per cent of $2\phi_s$. For instance, assume a constant ϕ' throughout the resetting half-cycle so that the flux will be reset by 4% of $2\phi_s$; then $\phi'_{(4\%)} = 0.04\ (4f\phi_s)$ and the average mmf requirement is

$$F_{4\%} = F_0 + a_1(0.96\phi_s) + a_3(0.16f\phi_s)^{a_3} \qquad (24)$$

The point thus calculated for Fig. 4 is in better agreement with the linearizations of the transfer characteristics.

For the calculation of the mmf F_{cq} which brings about the minimum or quiescent output either of two approximations appears feasible.

The one approximation considers the most critical requirement on E_c. Thus, for minimum output and for $E_g \leq E_{gN} = wN_g\phi_s$ the resetting core voltage essentially equals e_g/N throughout the reset half-cycle; then the peak value of $/\phi'/$ occurring at $3\pi/2$ is E_g/N_g, the value of flux is $\phi_s - E_g/wN_g$, and the mmf requirement at this instant may be evaluated from the core-reset function, equation 11. The control voltage required to produce these instantaneous values of F_c and ϕ' simultaneously may be evaluated from the control circuit voltage equation; the corresponding average control mmf F_{cq} is then

$$F_{cq} = F_0 + a_1\left(\phi_s - \frac{E_g}{\omega N_g}\right) + a_2\left(\frac{E_g}{N_g}\right)^{a_3} - \frac{E_g}{N_g}\,\frac{N_c^2}{R_c} \qquad (25)$$

The linearized core-reset function could also be used here. (In fact, in this particular case the 60-cps dynamic loop for sinusoidal voltage for proper value of E_g could be used; this is not true, however, in the following case.) For $E_g > E_{gN}$ minimum output results from flux reset to negative saturation. Then the core voltage is gate-limited during only a part of the resetting half-cycle, π to γ and δ to 2π, and is determined by the control circuit alone between γ and δ. Assuming that core voltage is constant between γ and δ, the most critical requirement on E_c occurs at δ. By equating the integral of core voltage in the resetting half cycle to $(-2\phi_s)$ γ, δ, and ϕ' in the interval γ to δ may be evaluated in terms of E_g/E_{gN}; a normalized curve for $\phi'(\gamma)$ is presented in Fig. 5. Assuming $\phi(\delta)$ is approximately zero, the value of F_{cq}, as determined by the critical requirement on E_c, may be found from Fig. 5 and the following equation:

$$F_{cq} = F_0 + a_2[-\phi'(\gamma)]^{a_3} + \phi'(\gamma)\frac{N_c^2}{R_c} \qquad (26)$$

Another possible means of approximating F_{cq} is to compute the average value of ϕ' over the reset half-cycle for the condition of minimum output and to use this value ϕ'_{avg} to approximate the critical value of E_c and the value of F_{cq}. Thus, for $E_g \leq E_{gN}$ and $\phi'_{avg} = 2E_g/(\pi N_g)$,

$$F_{cq} = F_0 + a_1\left(\phi_s - \frac{E_g}{\omega N_g}\right) + a_2\left(\frac{2E_g}{\pi N_g}\right)^{a_3} - \frac{2E_g}{\pi N_g}\,\frac{N_c^2}{R_c} \qquad (27)$$

and for $E_g \geq E_{gN}$ and $\phi'_{avg} = 4f\phi_s$,

$$F_{cq} = F_0 + a_2(4f\phi_s)^{a_3} - 4f\phi_s\frac{N_c^2}{R_c} \qquad (28)$$

Values of F_{cq} are plotted in Fig. 4 for normal gate voltage for each of these approximations. Agreement is best at large values of R_c/N_c^2; the first approximation gives conservative estimates of the slope of the characteristic whereas the second approximation yields optimistic values.

Each of these approximations, equations 25 through 28, indicates a linear relationship between F_{cq} and N_c^2/R_c; plots of the measured values of F_{cq}, and values of F_{cq} estimated by linearizing the experimental control characteristics, are both very close to a straight line, as shown in

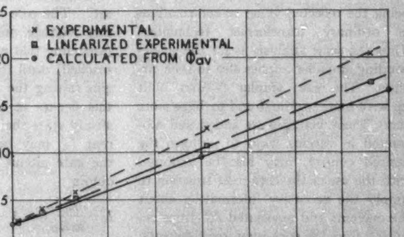


Fig. 6. Average control mmf for minimum output for various N_c^2/R_c, 2-winding half-wave amplifier, core 1A

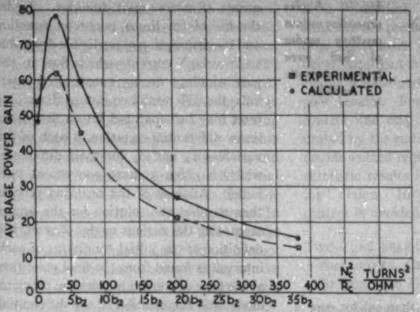


Fig. 7. Power gain versus N_c^2/R_c for 2-winding half-wave amplifier, core 1A

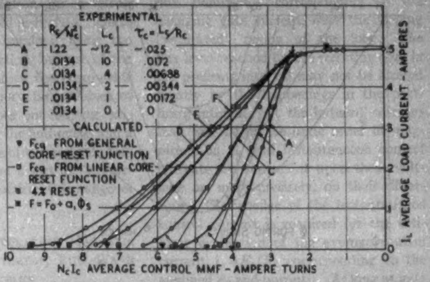


Fig. 8. Control characteristics of the 2-winding half-wave amplifier with inductive-control impedance at normal voltage, core 2A

Fig. 6. Values of F_{cq} calculated with the second of these approximations are plotted also, and lie within 9% of the values estimated by linearizing the control characteristics. Letting A_I represent the current gain of the amplifier or the slope $(dI_L)/(dI_c)$ over the linear portion of the control characteristic, a sort of average power gain $A_p = A_I^2\, R_L/R_c$ may be calculated; this is presented in Fig. 7 and compared with experimental values of A_p determined from the linearized control characteristic. The value of N_c^2/R_c for the optimum value of A_p is indicated fairly well by the calculations. Thus these approximations of the control characteristic have allowed quantitative predictions of the effects of control circuit resistance upon current and power gains and upon the shift of the control characteristic with control circuit resistance.

If inductive impedance is present in the control circuit of this half-wave amplifier it may be quite difficult to determine the various modes of operation simply by inspection. However, for $E_g \leq E_{gN}$ and $(R_c/N_c^2) >> (R_L/N_g^2)$ it is possible to determine critical control circuit conditions for minimum output and thereby ascertain a value for F_{cq}. At the minimum output level ϕ' is essentially equal to e_g/N_g throughout the cycle. Then the voltage equation of the control circuit

$$E_c = R_c i_c + L_c\frac{di_c}{dt} + N_c\phi' \tag{29}$$

may be solved for $F_c(t) = N_c i_c(t)$ to obtain

$$F_c = \frac{E_c/N_c}{R_c/N_c^2} + \frac{E_g/N_g}{R_c/N_c^2}\left(\frac{-\sin \omega t + \omega\tau_c \cos \omega t}{1+\omega^2\tau_c^2}\right) \tag{30}$$

where $\tau_c = L_c/R_c$. The condition that ϕ' just equals the gate voltage per turn can be met by insuring that i_L be greater than or equal to zero at all times so that the gate rectifier never blocks. More precisely, i_L must go to zero with a zero rate of change with respect to time; thus, the critical angle θ and the corresponding value of $F_{cq} = (E_c/N_c)/(R_c/N_c^2)$ may be found from the two conditions: $i_L(\theta) = 0$, $di_L(\theta)/dt = 0$. Approximating the core by the linear core-reset function, the gate circuit mmf is

$$N_g i_L = b_0 - F_{cq} + b_1\left(\phi_s - \frac{E_g}{\omega N_g}\right) + \frac{E_g}{N_g}\left(b_2 + \frac{N_c^2}{R_c(1+\omega^2\tau_c^2)}\right)\sin \omega t - \frac{E_g}{N_g}\left(\frac{b_1}{\omega} + \frac{N_c^2\omega\tau_c}{R_c(1+\omega^2\tau_c^2)}\right)\cos \omega t \tag{31}$$

and the first time derivative is

$$\frac{d(N_g i_L)}{dt} = \frac{\omega E_g}{N_g}\left[\left(b_2 + \frac{N_c^2}{R_c(1+\omega^2\tau_c^2)}\right)\times \cos \omega t + \left(\frac{b_1}{\omega} + \frac{N_c^2\omega\tau_c}{R_c(1+\omega^2\tau_c^2)}\right)\sin \omega t\right] \tag{32}$$

Setting this derivative equal to zero at $\omega t = \theta$ and solving for θ yields

$$\tan \theta = \frac{-\left(1 + b_2\dfrac{R_c}{N_c^2}(1+\omega^2\tau_c^2)\right)}{\dfrac{b_1}{\omega}\dfrac{R_c}{N_c^2}(1+\omega^2\tau_c^2) + \omega\tau_c} \tag{33}$$

and the corresponding value of F_{cq} for $N_g i_L = 0$ in equation 31 is

$$F_{cq} = b_0 + b_1\left(\phi_s - \frac{E_g}{\omega N_g}\right) - \frac{b_2 E_g}{N_g}\times \sqrt{\left(1 + \frac{1}{b_2\dfrac{R_c}{N_c^2}(1+\omega^2\tau_c^2)}\right)^2 + \left(\frac{b_1}{\omega b_2} + \frac{\omega\tau_c}{b_2\dfrac{R_c}{N_c^2}(1+\omega^2\tau_c^2)}\right)^2} \tag{34}$$

(A similar derivation can be performed with the use of the general core-reset function; the results are in slightly more complicated form.) Values of F_{cq} and $F_{4\%}$, found as described previously, are compared with experimental control characteristics in Fig. 8 for various values of τ_c and for normal voltage. The agreement is quite good inasmuch as each calculated value of F_{cq} lies close to or within the range of F_{cq}'s between the experimental and the linearized control characteristics.

HALF-WAVE BRIDGE MAGNETIC AMPLIFIER

The half-wave bridge magnetic amplifier[3] of Fig. 9 is a 2-core push-pull amplifier. In operation both cores are saturated at the end of the positive half-cycle of the alternating supply voltage and are reset in the negative half-cycle under the combined influence of the control mmf $N_c i_c$ and the bias mmf $N_b i_b$. The bias windings consist of only about 10% or 20% as many turns as the gate windings; the bias resistors R_b are needed

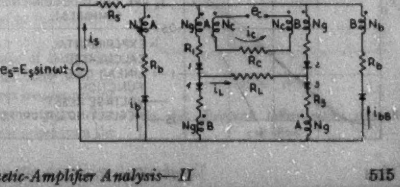


Fig. 9. Half-wave bridge magnetic amplifier

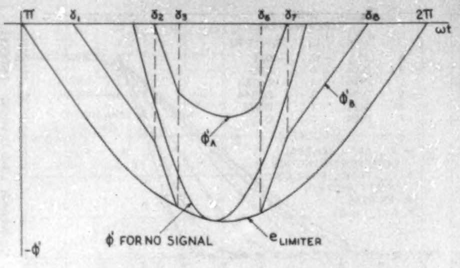


Fig. 10. Angles separating various resetting modes of half-wave bridge magnetic amplifier

to limit the bias currents. The control circuit also must have fairly high R/N^2 in order to limit the currents during the output interval. Thus, in calculating the reset of the cores, the nearly current-source behavior of the bias and the appreciable resistance of the control loop together make necessary a careful accounting of resistive drops; hence, an accurate knowledge of the core mmf requirements is needed.

During each resetting half-cycle there are several different modes of reset of the two cores. For positive signal voltage E_c the reset of core B under influence of bias and control begins at γ_1; core A begins reset later at γ_2; see Fig. 10. The resetting voltage of core B grows until it exceeds the limiting voltage of the source; then rectifiers 2 and 4 begin to conduct initiating an interval of gate-limited reset at γ_3 (γ_3 may occur before or after γ_2). Core A may also experience a gate-limited interval beginning at γ_4 and ending at γ_5, depending upon the bias mmf. The gate-limited reset of core B terminates at γ_6. After the end of the respective gate-limited intervals for each core, reset proceeds under the influence of control and bias until reset ends at γ_7 (which may be greater or smaller than γ_6) for core A, and at γ_8 for core B. In the positive half-cycle of supply voltage cores A and B both absorb voltage, with the two gate windings effectively in series until saturation of core A at α_A; output voltage then appears across the parallel combination of the resistor R_L and the two windings of core B until core B fires at α_B. After both cores have fired the current in the windings is limited primarily by the source resistance R_s. The cyclic average of output voltage is given approximately by

$$E_L = \frac{\omega N_g \Delta\phi}{2\pi\left(1+\frac{R_w}{R_L}\right)} \tag{35}$$

where R_w is the greater of R_1 and R_3, rectifier forward drops are neglected, and $\Delta\phi$ is the difference in preset fluxes of core A and core B and can be either positive or negative depending upon the sign of E_c.

Calculation of $\Delta\phi$ requires the solution the differential equations of reset for all the various modes. The differential equations for each interval are found by applying Kirchhoff's voltage law to the various loops and expressing core mmf by means of a core-reset function. Under the use of the linear core-reset function and the assumption of negligible drops due to "overlap" currents which flow in the gate windings during gate-limited intervals, the differential equations describing reset may be separated into a first-order linear differential equation in each of the variables ϕ_A and ϕ_B (or their difference, which is often a more convenient variable). Analysis of the resetting process then requires the solution for the angles separating the various modes of reset, the definition of the initial conditions of each interval as found from the final condition at the time of transition from the previous mode, and the solution of the differential equations of each interval. No attempt is made here to present a general analytical solution of this problem; a numerical solution was performed for a particular amplifier and the results are compared with the measured control characteristic in Fig. 11. Quite close agreement is obtained; note the poor agreement afforded under assumption of "voltage reset" neglecting i_cR_c.

Fast-Response Even-Harmonic Magnetic Amplifier

The fast-response even-harmonic magnetic amplifier[4,5] of Fig. 12 is a 2-core amplifier which in every half-cycle gives an output dependent upon the control during the same half-cycle; hence the term fast response. The half-cyclic integral of the alternating supply voltage is much larger than the sum of the swings of flux linkages of the N_s windings of the two cores, which are therefore driven into saturation in each half-cycle. At the end of the negative half-cycle of supply voltage both cores are at negative saturation. In the succeeding half-cycle both cores are reset toward positive saturation under the influence of the supply voltage and the control circuit; e_c is positive so that the voltage induced in the load circuit causes the rectifier to block as long as neither core is saturated. Core A reaches positive saturation first

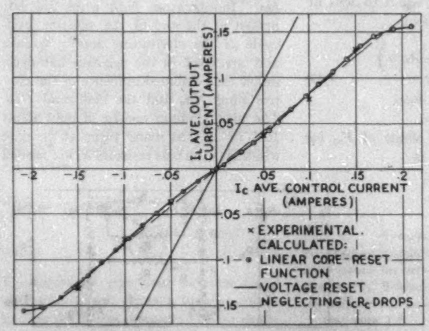
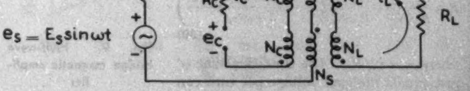


Fig. 11 (left). Control characteristic of half-wave bridge magnetic amplifier, cores 1A and 1B

Fig. 12 (below). Fast-response even-harmonic magnetic amplifier



at α_1; the remaining volt-time integral of core B then appears across the load resistor R_L. During the interval when load current flows R_c/N_c^2 is in parallel with R_L/N_L^2 across the unsaturated core; therefore, R_c/N_c^2 must be large with respect to R_L/N_L^2 to avoid loading the output or drawing excessive currents in the control loop. Core B fires at α_2 and both cores remain saturated until the end of the half-cycle; R_s serves to limit the source current during this interval. In the succeeding half-cycle the process is repeated; e_s has changed sign, e_c has not and thus the sequence of firing of the cores is reversed, with the rectifier still blocking when both cores are unsaturated. Only positive signals are used. The output is pulsating direct current of which either the d-c component or the associated harmonics may be used by the load.

The cyclic average of load current I_L is given in terms of the difference $\Delta\phi$ of the two core fluxes at the first firing angle α_1 by

$$I_L = \frac{\omega N_L \Delta\phi - E_f(\alpha_2 - \alpha_1)}{\pi R_L} \tag{36}$$

where α_2 is the later firing angle. Since the control resistance is necessarily different from zero, $i_c R_c$ drops during reset may well be appreciable. The core voltages are large during reset so the linear core-reset function is applicable. Core A begins resetting at γ_1, core B at γ_2, and firings occur at α_1 and α_2 respectively. Here again, the calculation of $\Delta\phi$ requires the solution of the differential equations for the various intervals, calculation of the angles separating the various modes, and evaluation of the boundary conditions. The results of a numerical solution for a given amplifier are presented in Fig. 13. Quite close agreement is obtained using the linear core-reset function, but use of the 3-zone approximation of the core yields poor results. For very large values of control one core does not become saturated by the end of the half-cycle, and a double-valued region of the control characteristic results from a sort of "triggering" action. In practice, this range of control is avoided.

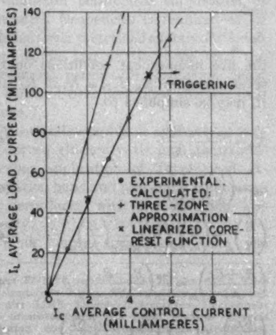


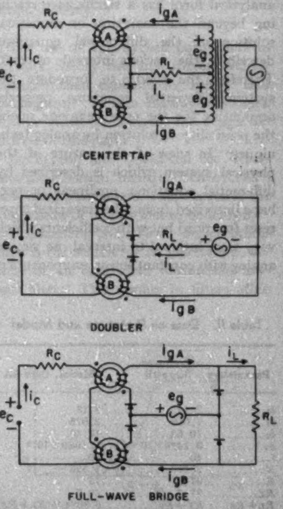
Fig. 13. Control characteristic of fast-response even-harmonic magnetic amplifier, cores 1A and 1B

Two-Core Self-Saturating Magnetic Amplifiers, Centertap, Full-Wave Bridge, and Doubler

The analysis of conventional 2-core self-saturating amplifiers, centertap, full-wave bridge, and doubler, Fig. 14, controlled from a constant current source, follows as an immediate extension of the analysis of the single-core half-wave amplifier with constant control current. In the 2-core amplifiers the average value of the load current may be easily evaluated in terms of the preset flux and gate voltage as the sum of the average gate currents which flow in alternate half-cycles. However, as one core resets, the value of the limiter voltage in its gate loop depends upon the gate current of the other core due to the common load voltage. Thus, operation of the 2-core amplifiers is not identical with that of two independent half-wave amplifiers, but the reset can be calculated with a proper evaluation of the limiter voltages in the gate loops. Results of such calculations are presented in Fig. 15 for a centertap amplifier; fair agreement is obtained with the experimental characteristic taken at the comparatively large value of R_c/N_c^2.

If the condition of constant control current is removed, an electromagnetic coupling between the cores through the control loop also becomes effective. At the beginning of a half-cycle, one core is in saturation and the other is at the preset flux level ϕ_p. In the prefiring interval the latter core absorbs the gate voltage until it saturates at α; meanwhile, the previously saturated core resets under the combined action of the control voltage E_c, the resistive drops $R_c i_c$ of the control circuit, and the voltage induced in the control circuit by the changing flux of the gating core. The core-reset function may be used in this interval to calculate ϕ and ϕ' in the resetting core.

After firing, the voltage of the gating core is nearly zero and may well be ignored at ordinary power frequencies for purposes of evaluating outputs. The voltage which does appear arises from the flux excursions, corresponding to the changing gate current, over the sloping "saturated" region of the loop. The flux level reached by the resetting core in the prefiring interval may not be altered greatly by the happenings in the postfiring interval, but the induced voltages may well be of the same order of magnitude as E_c and instantaneous control currents are dependent upon these voltages or, more precisely, on their difference. (Knowledge of the instantaneous control current is required for the ultimate evaluation of the average control current $I_c = E_c/R_c$ corresponding to the assumed ϕ_p and output.) As long as gate current continues to grow, the flux of the gating core increases, and the voltage induced in the control circuit remains in a direction to reset the other core, although the rate is much slower than before firing. But as the gate current recedes from its maximum value after the middle of the half-cycle, the induced voltage changes sign, opposing E_c, in a direction to return the reset core toward saturation even before the end of the half-cycle. Thus, for sinusoidal gate voltages, the instantaneous control current after firing cannot be evaluated under the use of the core-reset function, for the magnetic processes involve flux changes along the "top" and "bottom" and ascending branch of the minor loops where the



CENTERTAP

DOUBLER

FULL-WAVE BRIDGE

Fig. 14. Two-core self-saturating magnetic amplifiers

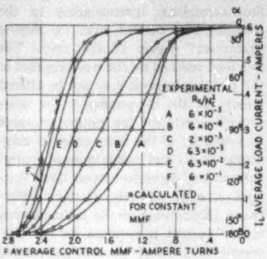


Fig. 15. Control characteristic of centertap amplifier with various R_c/N_c^2 and normal voltage, cores 1A and 1B

core-reset function is not applicable. However, for a square-wave gate voltage, the postfiring gate current would be essentially constant; the troublesome voltages induced in the control winding of the saturated core after firing would be eliminated, and the core-reset function could be used throughout the half-cycle to describe core reset.

Model Theory for Design of Analog Computers for Magnetic-Amplifier Systems

The availability of a core-reset function expressing dynamic core-reset in analytical form has a significance reaching beyond the analytical or numerical solution of the differential equations describing the various intervals of reset. Therefore the ability to formulate the system differential equations, properly accounting for the core behavior, opens the possibility of solution by analog techniques. In view of the nature of the physical system, which is described by differential equations, nonlinear, or perhaps linearized (using the linearized core-reset function) but with coefficients which vary from interval to interval, no simple analog with constant linear components is suitable. Electronic means could conceivably be used to synthesize an appropriate analog, but it appears more convenient to use magnetic cores as the basic components. The question is then reduced to one of formulating the rules of similitude governing the analogy between the system whose performance is to be predicted (the prototype) and the system on which measurements are to be performed (the model).

A comprehensive model theory aiming at predictions of behavior of a prototype on the basis of pertinent measurements performed on some different amplifier available for experimentation has been developed.[6] The rules presented, however, have been developed under the assumption of an existing affinity of the ϕ, F relationship (single- or double-valued) of the model and prototype, while the more recently acquired appreciation of dynamic phenomena (the influence of ϕ') was not considered. Analogies established under use of the more complete core-reset functions are, of course, much more exacting in the choice of the model cores, in the determination of the remaining components, and in the specification of the significant tests to be performed on the model. It appears likely that in some instances the problem of synthesis may be overdetermined and therefore impossible, but in other instances this analog approach may be fruitful.

As an example of the use of the model theory, consider the single-winding single-core half-wave amplifier, Fig. 1(A), controlled by a resistor R_s shunting the rectifier. In general, two physical systems are analogous only if they are described by the same set of differential equations with the same boundary conditions. To meet the restrictions on the core-reset function, the flux must reset from positive saturation; this restriction is met by the physical situation, and it will be convenient to measure the flux in terms of the difference from saturation: $\lambda=\phi-\phi_s$ and $\lambda'=\phi'$. The rectifier is assumed to act as an ideal switch. The mmf requirement of the core could be written in terms of the general core-reset function, but this would introduce the term $(-\lambda')^{a_3}$; a_3 is a dimensionless constant characteristic of the system, and one requirement will surely be that $a_{3m}=a_{3p}$, where subscripts m and p refer respectively to model and prototype. This condition may not necessarily imply an overdetermination of the conditions of analogy, but it may be difficult to meet practically in the selection of cores for the model. Thus it will be more convenient to develop rules of similitude under use of the linear core-reset function, rewritten in terms of λ as

$$F=b_0+b_1\phi_s+b_1\lambda+b_2\lambda' \tag{37}$$

where $0\geq\lambda\geq-2\phi_s$. To write the differential equations in dimensionless form the variable $\theta=\omega t$ is introduced; then $\lambda'=d\lambda/d\theta\times d\theta/dt=\omega d\lambda/d\theta$. Making these substitutions causes the differential equation of reset to become

$$\left[1+b_2\left(\frac{R_s+R_L}{N^2}\right)\right]\frac{\omega N}{E}\,\frac{d\lambda}{d\theta}+\frac{(R_s+R_L)/N^2}{E/N}\times$$
$$b_1\lambda+\frac{(R_s+R_L)/N^2}{E/N}\,(b_0+b_1\phi_s)=\sin\theta \tag{38}$$

In the gating half-cycle the differential equation is

$$\frac{\omega N}{E}\,\frac{d\lambda}{d\theta}+\frac{R_L/N^2}{E/N}\,F=\sin\theta \tag{39}$$

since the core-reset function does not apply for ascending fluxes, no substitution is made here for F. Relating fluxes and mmf of model and prototype as follows:

$$\lambda_p=K_\phi\lambda_m \tag{40}$$

and

$$F_p=K_fF_m \tag{41}$$

the five independent conditions on the seven variables: K_ϕ, K_f, ω, N, E, R_L, and R_s may be simplified to

$$K_\phi=\frac{(\phi_s+b_0/b_1)_p}{(\phi_s+b_0/b_1)_m} \tag{42}$$

$$\omega_m=\left(\frac{b_2}{b_1}\right)_p\left(\frac{b_1}{b_2}\right)_m\omega_p \tag{43}$$

$$\left(\frac{E}{N}\right)_m=\left(\frac{b_2}{b_1}\right)_p\left(\frac{b_1}{b_2}\right)_m\frac{1}{K_\phi}\left(\frac{E}{N}\right)_p \tag{44}$$

$$\left(\frac{R_s+R_L}{N^2}\right)_m=\frac{b_{2p}}{b_{2m}}\left(\frac{R_s+R_L}{N^2}\right)_p \tag{45}$$

$$\left(\frac{R_L}{N^2}\right)_m=\left(\frac{b_2}{b_1}\right)_p\left(\frac{b_1}{b_2}\right)_m\frac{K_f}{K_\phi}\left(\frac{R_L}{N^2}\right)_p \tag{46}$$

Two further relationships between E, N, R_S, R_L, and K_f may be imposed;

Table II. Data on Prototype and Model

Parameters	Prototype, Core 1B	Model, Core 2A
b_0	2.14	3.13
b_1	1,059	2,075
b_2	10.62	10.40
ϕ_s	0.1478×10^{-3}	0.1965×10^{-3}
f	40	80
N	2,776	1,523
E	97.5	75
R_L	77.7	59.7
R_L+R_S	R_L+R_S	$0.3074\times(R_L+R_S)$
I	I	I
K_ϕ	1.428	
K_f	1.823	

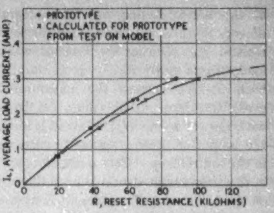


Fig. 16. Control characteristic of prototype and prediction from model

ω_m and K_a are fixed by the parameters of the prototype and the properties of the core used for the model.

An experimental check of this analysis was made with the use of cores and circuit parameters as listed in Table II. In Fig. 16 the control characteristic calculated for the prototype from measurements on the model is compared with the actual experimental data for the prototype. The agreement is fairly good, the error in predicted values being less than 10%. Sources of error aside from the use of the linearized core-reset function may be in the rectifier imperfections (leakage and forward drop), and the nonzero slope of the F, ϕ relation in the saturated region.

This is but one example of the solution of magnetic-amplifier problems by analog methods. No further attempt will be made here to investigate the ramifications of the model theory or applications to other circuits or systems, but it is suggested that the improved representation and understanding of dynamic core behavior opens the field for devising and performing pertinent experimentation on analogous circuits to determine steady-state and transient performance of magnetic amplifiers. There may be limitations as to the circuits which may be treated in this way, but it is encouraging to note that for the foregoing example the analogous system was not overdetermined and that there was some freedom of choice in selecting the parameters of the model.

Summary

The core-reset functions which were developed in Part I to better express core behavior under dynamic conditions of changing flux have been used here in several examples to calculate the steady-state magnetic-amplifier performance with quite good accuracy. Although a thorough analysis becomes somewhat involved, the computational work may be readily performed with digital computers. Moreover, the core-reset functions lead to some results of general applicability in magnetic-amplifier analysis and design, and the understanding of simple reset problems and the approximate solution of more complex problems are both aided by the improved description of core behavior. The possible use of magnetic cores as analog computer components for the calculation of steady-state and transient performance of magnetic-amplifier systems is suggested as a potent analytical tool to be developed further.

References

1. CORE-RESET FUNCTIONS IN MAGNETIC AMPLIFIER ANALYSIS, PART I—A CORE-RESET FUNCTION, G. C. Feth. *AIEE Transactions*, vol. 77, pt. I, Sept. 1958, pp. 503–09.

2. THE OPERATION OF MAGNETIC AMPLIFIERS WITH VARIOUS TYPES OF LOADS, PART I—LOAD CURRENTS FOR GIVEN ANGLE OF FIRING, L. A. Finzi, R. R. Jackson. *Ibid.*, vol. 73, pt. I, July 1954, pp. 270–78.

3. AN IMPROVED MAGNETIC SERVO AMPLIFIER, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 281–89.

4. FAST-RESPONSE MAGNETIC SERVO AMPLIFIER, Frank Hill, Joseph A. Fingerett. *Electronics*, New York, N. Y., Oct. 1954.

5. THE STUDY OF A MAGNETIC INVERTER FOR AMPLIFICATION OF LOW-INPUT-POWER D.C. SIGNALS, E. H. Frost-Smith. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. II, 1953, p. 362.

6. RULES OF SIMILITUDE FOR MAGNETIC AMPLIFIER SYSTEMS, L. A. Finzi, H. Durand. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 7, Feb. 1952, pp. 498–514.

Insertion Loss Filter Design Applied to Transistorized Carrier System

T. WINKLER
NONMEMBER AIEE

IN MANY filter design problems, the minimum permissible suppression of unwanted frequencies is required to be constant over prescribed effective attenuation bands, and at the same time the maximum permissible distortion must be constant over prescribed effective pass bands.

Most design problems of this type may be solved efficiently through the simultaneous application of Tchebycheff attenuation band and pass-band parameters in accordance with the insertion loss theory.

While more complicated than the image-parameter theory, which has dominated practical filter design almost exclusively for many years, this theory sometimes permits the realization of greater network economy exhibiting superior and better controlled electrical characteristics with no increase in cost. Its general problem is to find a reactance 4-pole yielding prescribed insertion loss versus frequency function when inserted between prescribed resistance terminations.

Although a general application of the insertion loss method is theoretically possible, complicated numerical computations requiring a very high accuracy originally limited its practical usefulness. The introduction, however, of computing equipment, that has become in recent years a steady companion of a modern electronics designer, and a growing tendency toward miniaturization (especially after the invention of the transistor) combined with performance requirements, not easily satisfied with image-parameter designs, have led to an extensive use of the insertion loss theory in recent years.

Basic Theory

The use of Tchebycheff parameters in network theory was first introduced by Cauer,[1,2] then explored by Piloty[3] and Darlington.[4] It was Darlington's work that provided the basis for the design described in this paper.

Darlington also originated the "reference filter" design, based on Tchebycheff's pass-band parameters, which requires the frequencies of infinite loss to be known.

The starting point of the design procedure is the search for an insertion power ratio of an equivalent normalized low-pass filter. This applies to filters other

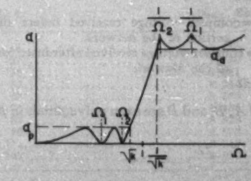


Fig. 1. A characteristic of a normalized low-pass filter, for $\eta=2$

Paper 57-1070, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted May 27, 1957; made available for printing April 25, 1958.

T. WINKLER is with the Stromberg-Carlson Company, Rochester, N. Y.

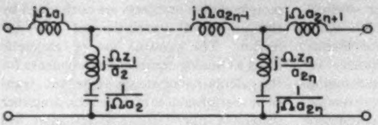

$j\Omega a_1$
$j\Omega a_{2n-1}$
$j\Omega a_{2n+1}$
$j\frac{\Omega z_1}{a_2}$
$\frac{1}{j\Omega a_2}$
$j\frac{\Omega z_n}{a_{2n}}$
$\frac{1}{j\Omega a_{2n}}$

Fig. 2. Impedance branches of a mid-series low-pass ladder

than low-pass filters. In the remaining case the actual filter takes the place of the equivalent structure, except for normalization of its impedance and frequency scale.

The insertion loss versus frequency characteristic of the normalized low-pass filter is shown in Fig. 1, in which η represents the number of "sections." Its value depends on α_p, α_s and k. Since η must be an integer, one of these variables may have to be modified.

Darlington's corresponding power ratio, required for a symmetrical low-pass filter, takes the form

$$\epsilon^{2\alpha}=1+\left[S_0\frac{\Omega(\Omega^2-\Omega_1^2)\ldots(\Omega^2-\Omega_n^2)}{(1-\Omega_1^2\Omega^2)\ldots(1-\Omega_n^2\Omega^2)}\right]^2 \quad (1)$$

where

$$\Omega=\frac{\omega}{\sqrt{\omega_1\omega_2}} \quad (2)$$

with ω_1 and ω_2 being the values of ω at cutoff frequency and the limit of the effective attenuation band,

$$\Omega_\sigma=\sqrt{\kappa}\ \mathrm{sn}\left(\frac{2\sigma K}{2\eta+1}, \kappa\right) \qquad \sigma=1, \ldots, \eta \quad (3)$$

where

sn = elliptic sine

$\kappa=\omega_1/\omega_2$ = modulus

$K=\int_0^{90^\circ}\frac{d\varphi}{\sqrt{1-\kappa\sin^2\varphi}}$ = "complete elliptic" integral of the first kind

S_0 = function of α_p in Fig. 1

After the roots of equation 1 have been determined, two ways of realization are available. The first is by the corresponding complex insertion voltage ratio

$$\frac{V_{20}}{V_2}=\frac{A+pB}{P} \quad (4)$$

where

V_{20} = complex voltage received before insertion of the network

V_2 = complex voltage received after insertion of the network

$p=j\Omega$

and A, B, and P are even polynomials in p, with real coefficients that may be determined by making use of the definition

$$\epsilon^{2\alpha}=\left|\frac{V_{20}}{V_2}\right|^2 \quad (5)$$

In addition to polynomials A and B, the polynomial

$$B'=-\Pi_\sigma(p^2+\Omega_\sigma^2) \quad (6)$$

is evaluated.

A, B, and B' permit the determination of the short-circuit driving point impedance of the normalized low-pass structure

$$Z_{s1}=\frac{p(B-B')}{A} \quad (7)$$

The continued fraction expansion of

$$F=\frac{Z_{s1}}{p}=a_1+\cfrac{1}{\cfrac{a_2}{z_1-z}+\cfrac{1}{a_3+\ldots}} \quad (8)$$

with

$$z=-\frac{1}{p^2}=\frac{1}{\Omega^2}$$

$$z_v=\Omega_v^2 \qquad v=1, \ldots, \eta \quad (9)$$

supplies the ladder-network coefficients $a_1, \ldots, a_{2n+1}$, whose meaning is explained in Fig. 2. The second realization method is by using the insertion power ratio expression

$$\epsilon^{2\alpha}=1-\left(\frac{pB'}{P}\right)^2 \quad (10)$$

The quantity

$$P+pB'=(P_1+pB_1)(P_2-pB_2) \quad (11)$$

could be determined by associating all roots having negative real parts with (P_1+pB_1). With the use of the identities

$$Z_A=R\frac{pB_2}{P_2},\ Z_B=R\frac{P_1}{pB_1} \quad (12)$$

one lattice, producing the prescribed power ratio, could be built (Fig. 3) and then transferred into an equivalent ladder of the kind shown in Fig. 2.

It turns out that there are four corresponding lattices. The one chosen conforms with the already available table of special design formulas.[4]

Although the methods described apply to symmetrical networks only, the first could be modified easily to permit the realization of inverse impedance networks.

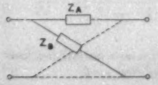


Fig. 3. The lattice configuration

The second method has been used in the design described in the following.

Design Objectives and Procedures

The attenuation requirements for the *561* subscriber carrier channel band-pass filters are given in Fig. 4.

Ten carrier frequencies have been assigned between 15 and 123 kc (kilocycles per second) in equal steps of 12 kc. The filters should meet the requirements shown in Fig. 4 when all ten are operated in parallel.

It has been mentioned that the theory and procedures of low-pass filter design also include the cases of band-pass filters.

Where the band-pass insertion loss specification has complete geometric symmetry about some mid-band frequency, the well-known band-pass/low-pass transformation is applicable (see chap. 10 of reference 5). After designing a low-pass filter to this transformed specification the band-pass filter is arrived at by a simple frequency transformation of the network elements themselves, in which inductances and capacitances are replaced by series- and parallel-resonant circuits respectively.

Although the foregoing requirements do not contain the necessary symmetry, this method of design, producing higher-than-required attenuation in the lower frequency transition region, is less complicated and supplies more design flexibility than any known design procedure applicable to structures with other than geometric symmetrical attenuation.

Darlington[4] and Boothroyd[6] have independent ways of solving such problems; the former's is based on Norton's impedance transformation and latter's on low-pass filter design with different attenuations for positive and negative frequencies. The author's treatment of this problem with the aid of an electronic computer and application of a conformal transformation is scheduled to be published.

The maximum pass-band loss of 1.8 db

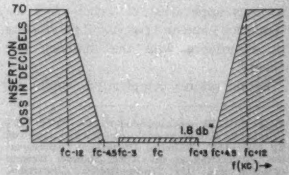


Fig. 4. Insertion loss specification

*The insertion loss between f_c-3 and f_c+3 should be flat within 25 db

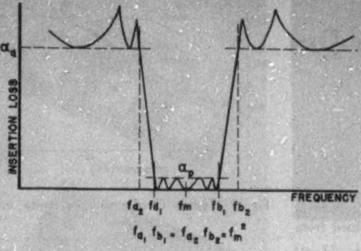


Fig. 5. Form of filter characteristic meeting the design specifications

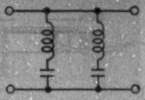


Fig. 7. Alternative shunt branch configuration

(decibels) required a design on a nondissipative basis. The effect, however, of the parasitic dissipation, which must be present in an actual reactance network, cannot be neglected and its effect on the theoretical attenuation curve should be predicted and accounted for. The form of insertion loss characteristic of a band-pass filter described in the foregoing is shown in Fig. 5.

Should the parasitic dissipation be neglected, all constants, but one, required to establish the corresponding low-pass power ratio (equation 1) are as follows:

$$k=\frac{f_{b1}-f_{a1}}{f_{b2}-f_{a2}}$$

$$\alpha_p=0.3\times 0.11513=0.034539 \text{ neper}$$

$$\alpha_a \geqq 70 \text{ db} \qquad (13)$$

In this case the frequencies f_{a1}, f_{b1}, and f_{b2} correspond to f_{c-3}, f_{c+3}, and f_{c+12} in Fig. 4 respectively; $f_{a2}=f_{a1}\,f_{b1}/f_{b2}$ then falls into the transition region.

The missing constant required in equation 3 is η, which is the same for both filters provided sections of the 6-element type of configuration are considered in the band-pass case. Its minimum value is determined from its relation with α_p, α_a, and k, given by the equation

$$\alpha_a=[10\log_{10}(e^{2\alpha_p}-1)-10(2\eta+1)\log_{10} q-12.04]\text{db} \qquad (14)$$

in which $\log_{10}(q)$ is a quantity which is tabulated against $\sin^{-1}k$ in elliptic function tables (see page 49 of reference 7).

Equation 14 is an approximation. It follows from the equation

$$e^{2\alpha_a}=1+\frac{e^{2\alpha_p}-1}{\kappa_1^2}$$

which for large α_a and small α_p reduces to

$$e^{2\alpha_a}\approx\frac{e^{2\alpha_p}-1}{\kappa_1^2}\approx\frac{e^{2\alpha_p}-1}{16q_1\dfrac{1+4q_1^2}{1+8q_1+24q_1^2}}\approx\frac{e^{2\alpha_p}-1}{16q_1}$$

and with use of the identity $q_1=q^{2\eta+1}$ assumes the required form

$$e^{2\alpha_a}\approx\frac{e^{2\alpha_p}-1}{16q^{2\eta+1}}$$

Using equation 14 and remembering than η is the number of ladder sections, a table of values for various k's, with $\eta=1,2,3,\ldots$ as a parameter, could be computed. This has been done for the lowest frequency network with the following results:

η	α_a, db
1	41.98
2	84.63
3	129.29

with

$$k=\frac{6}{27-\dfrac{12\times 18}{27}}=0.315789 \qquad (15)$$

and

$$\alpha_p=0.3 \text{ db}$$

Use of two sections seemed to be required. The excess of attenuation, however, had to be exchanged for smaller α_p and larger k, to provide enough margin for the effect of parasitic dissipation, manufacturing variations, and temperature effect.

After choosing $\alpha_p=0.1$ db and $\sin^{-1}k=22$ degrees, equation 14 establishes α_a to be 72.8 db.

The next step is the evaluation of bandwidth $D=f_{b1}-f_{a1}$, corresponding to k. As a result, the required equation takes the form

$$D_{\max}=\left[\frac{2(f_c+12)}{k}-\sqrt{\left[\frac{2(f_c+12)}{k}\right]^2-96(f_c+6)}\right]\text{kc} \qquad (16)$$

with f_c being the carrier frequency called for in Fig. 4, expressed in kc.

Equation 16 reveals another very important aspect of this procedure: with the sharpness of cutoff, expressed in the value of k, being constant, the theoretical maximum bandwidth $D_{\max}$ will increase when the carrier frequency f_c increases. The values of $D_{\max}$ corresponding to the prescribed f_c's have been computed as:

f_c, Kc	$D_{\max}$, Kc
15	7.179
27	7.760
39	8.061
51	8.244
63	8.368
75	8.456
87	8.523
99	8.576
111	8.618
123	8.652 (17)

This behavior of $D_{\max}$ permitted the use of a single equivalent low-pass filter,

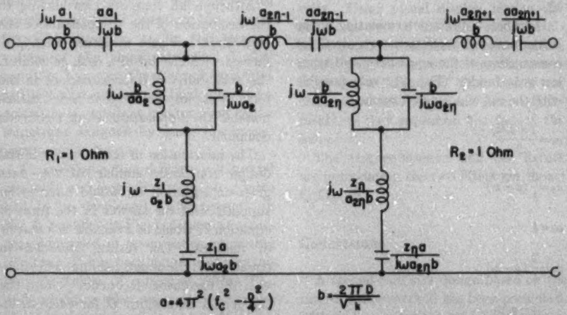

$j\omega\frac{a_1}{b}$
$\frac{aa_1}{j\omega b}$
$j\omega\frac{a_{2\eta-1}}{b}$
$\frac{aa_{2\eta-1}}{j\omega b}$
$j\omega\frac{a_{2\eta+1}}{b}$
$\frac{aa_{2\eta+1}}{j\omega b}$
$j\omega\frac{b}{aa_2}$
$\frac{b}{j\omega a_2}$
$j\omega\frac{b}{aa_{2\eta}}$
$\frac{b}{j\omega a_{2\eta}}$
$R_1=1$ Ohm
$R_2=1$ Ohm
$j\omega\frac{z_1}{a_2 b}$
$j\omega\frac{z_\eta}{a_{2\eta} b}$
$\frac{z_1 a}{j\omega a_2 b}$
$\frac{z_\eta a}{j\omega a_{2\eta} b}$
$a=4\pi^2\left(f_c^2-\frac{D^2}{4}\right)$
$b=\frac{2\pi D}{\sqrt{k}}$

Fig. 6. Impedance branches of a band-pass filter derived from the preliminary low-pass filter (Fig. 2) by a frequency transformation; inductances are in henrys, capacities in farads

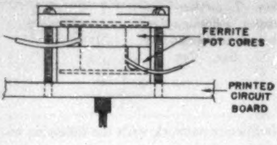


Fig. 8. Variable inductor

whose characteristic is plotted against the normalized frequency in Fig. 1.

The required design constants have been chosen to be

$$k=0.37461$$

$$\eta=2$$

$$\alpha_p=0.1 \text{ db} \tag{18}$$

Using the approximation

$$p_0=\frac{\sqrt{k}\,\mathrm{sn}\,(u',\sqrt{1-k^2})}{\mathrm{cn}\,(u',\sqrt{1-k^2})} \tag{19}$$

where

$$u'=\frac{2K}{\pi(2\eta+1)}\sinh^{-1}\frac{1}{\sqrt{\epsilon^{2\alpha_p}-1}}$$

the evaluation of the real root of the power ratio equation 1 was possible. The error in p_0 will produce only a small change in the original value of α_p and a respective small change in α_a in accordance with equation 14.

The expression

$$\mathrm{sn}\,(u',\sqrt{1-k^2})=\mathrm{sn}\,(rK',k') \tag{20}$$

with

$$r=u'/K'$$

$$k'=\sqrt{1-k^2}$$

K' = complete integral corresponding to modulus k'

was evaluated using the approximation[8]

$$\mathrm{sn}\,(rK',k')=\frac{1}{\sqrt{k'}}\frac{(1-t^r-t^{2-r}+t^{2+2r}+\dots)}{(1+t^r+t^{2-r}+t^{2+2r}+\dots)} \tag{21}$$

where

$$t=\frac{1}{2}\frac{1-\sqrt{k'}}{1+\sqrt{k'}}+\dots$$

The values of K are tabulated against $\sin^{-1}k$ in elliptic function tables (see page 85 of reference 7) and in addition there are the following rapidly convergent series:

$k^2 \ll 1$:

$$K=\frac{\pi}{2}\left(1+2\Delta+9\Delta^2+50\Delta^3+\frac{1{,}225}{4}\Delta^4+\dots\right)$$

with

$$\Delta=k^2/8$$

$k^2 \approx 1$:

$$K=\Lambda+\frac{\Lambda-1}{4}k'^2+\frac{9}{64}\left(\Lambda-\frac{7}{6}\right)k'^4+\frac{25}{256}\left(\Lambda-\frac{37}{30}\right)k'^6+\dots$$

with

$$k'=\sqrt{1-k^2}$$

$$\Lambda=\ln(4/k')$$

After sn (u',k') has been determined, the value of cn (u',k') is obtained from the equation

$$\mathrm{sn}^2+\mathrm{cn}^2=1 \tag{22}$$

Using equations 19 through 21 the positive real root is found to be

$$p_0=0.34627351 \tag{23}$$

The complex roots are then evaluated using the approximation equation 21 and the formula

$$p_\sigma=\frac{(-1)^\sigma p_0 cd_\sigma \pm jW\Omega_\sigma}{1+p_0^2\Omega_\sigma^2}=(-1)^\sigma{}_{\gamma\sigma}\pm j\beta_\sigma,\quad \sigma=1,2 \tag{24}$$

where

$$W=\sqrt{(1+kp_0^2)\left(1+\frac{1}{k}p_0^2\right)}$$

$$cd_\sigma=\sqrt{(1-\mathrm{sn}_\sigma^2)(1-k^2\mathrm{sn}_\sigma^2)}$$

$$\Omega_\sigma=\sqrt{k}\,\mathrm{sn}_\sigma$$

$$\mathrm{sn}_\sigma=\mathrm{sn}\left(\frac{2\sigma}{5}K,k\right)$$

Their values are

$$p_1=-0.26526433\pm j0.42532405$$

$$p_2=0.09338782\pm j0.65868852 \tag{25}$$

After the roots of the power ratio have been computed, the impedance arms of the corresponding lattice are easily available with the aid of equations 11 and 12.

However, a direct determination of the ladder-network coefficients, required for construction of the equivalent mid-series low-pass ladder (Fig. 2), was possible with the aid of a set of formulas

$$a_1=\frac{1-\rho_1^2z_1}{2\gamma_1}$$

$$a_2=\frac{2}{\delta}(\psi-z_1)$$

$$a_3=\delta$$

$$a_4=\frac{2}{\delta}(\psi-z_2)$$

$$a_5=\frac{1-\rho_1^2z_2}{2\gamma_1} \tag{26}$$

where

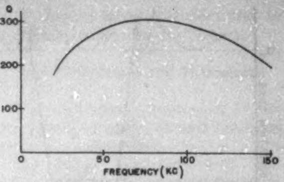


Fig. 9. Typical Q of a variable ferrite inductor

$$\delta=\frac{\rho_1^2}{\gamma_1}\left(\psi-\frac{z_1+z_2}{2}\right)$$

$$\psi=\frac{p_0+2_{\gamma_2}}{p_0\rho_2^2}$$

$$\rho_r^2=\gamma_r^2+\beta_r^2$$

These formulas were computed using equations 11 and 12 and Darlington's Table II.[4]

The coefficients could also be evaluated from equations 8 and 9 by replacing Z_{s1} with $Z_{01}=1/2(Z_A+Z_B)$, as

$$F=\frac{Z_{01}}{p}=\frac{1}{2}\left(\frac{Z_A}{p}+\frac{Z_B}{p}\right)=F_1(z)=\frac{1}{2}(F_{1A}+F_{1B})$$

and by using the formulas given in the Appendix.

The coefficient $a_{2\eta+1}$, corresponding to the inductance forming the last series arm, may then be obtained from the formulas for a_1, by replacing z_1 with z_η.

By combining equation 9, 24, and 26 the numerical values are determined as:

$$z_1=0.13544096$$
$$z_2=0.34077093$$
$$a_1=1.82076559$$
$$a_2=2.17832867$$
$$a_3=3.06903102$$
$$a_4=2.04452098$$
$$a_5=1.72351847 \tag{27}$$

Although all computations leading to the evaluation of the components of the equivalent low-pass structure were performed with the aid of a desk calculator, the availability of approximations in the form of rapidly convergent series makes possible the application of an electronic computer.

The information in terms of the initial design constants, similar to the ones given in equation 18, would have to be supplied, and the answer in the form of equation 27 would be available in a matter of minutes. The results obtained in equation 27, the design constant k in equation 18, the carrier frequency f_c, and the theoretical bandwidth D for each structure are required for the evaluation of components of the respective band-pass filter, using the impedance formulas given

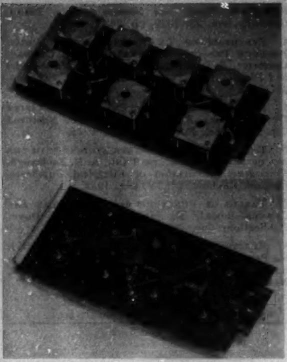


Fig. 10. Typical filter assembly

in Fig. 6. The shunt branches are replaced by configurations of the type shown in Fig. 7.

The impedance level was not originally defined because of the flexibility of the assigned transformation ratio of the mutual repeating coil; therefore, the terminations R_1 and R_2, representing the generator internal resistance and the load respectively, were chosen to be 300 ohms.

At this level the coils in all ten filters assumed values permitting the resistance-reactance ratios $d_L = 1/Q_L$ to be less than 1/200 within the respective pass bands.

At the same time, the required capacitors, with the exception of very few in lower frequency units, were in the mica range.

Physical Realization

The distinguishing feature of the insertion loss theory is the use of an exactly prescribed insertion loss versus frequency function to fix the element values of the final network. Unfortunately, however, the effect of variations in these elements on the final filter characteristic is somehow greater than in the structures designed by using the image-parameter technique.

To overcome this deficiency a special assembly and new adjustment procedures had to be developed. The network in Fig. 6, modified in Fig. 7, presents a structure composed of seven series-resonating circuits. It has been found that controlling the values of capacities within 1%, and adjusting the resonant frequencies of each circuit to their nominal values, would produce adequate results.

While 1% capacitors were easily available, variable inductors were required to make the frequency adjustments possible and practical. An inductor of this kind is shown in Fig. 8.

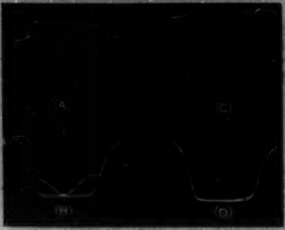


Fig. 11. Loss characteristic of the standard filter and a filter under test

A, B—Attenuation bands and pass band before adjustment
C, D—After adjustment

Two ferrite pot core halves, with center posts separated by an air gap, a ferrite slug whose position within the air gap may be adjusted with a screw driver, and a plastic bobbin are the basic elements of the assembly.

The dimensions were chosen to make possible manufacture of all inductors used in these filters with only two standard air gaps. While their inductance range was somewhere around ±10%, requiring a different number of turns for most of the specified values, the Q and the temperature coefficient of each coil met the requirements with considerable safety margin. A typical Q versus frequency performance is shown in Fig. 9 and a filter assembly in Fig. 10.

Use of adjustable inductors also offered great manufacturing advantages. Only one filter of a kind had to be carefully adjusted in accordance with the design results. Using visual display testing circuits, the rest of the units could be almost completely adjusted on a comparison basis. Fig. 11 shows the loss characteristics of the "standard" filter and of one filter under test before and after adjustments, as they appear on the face of the scope.

The test equipment and the fixture accommodating the two filters are shown in Fig. 12.

Conclusions

A special case of a design based on the insertion loss method has been presented. Although a complete set of formulas is given for a design of symmetrical band-pass filters with two 6-element sections, design of structures containing a different number of sections η (corresponding to a specific set of requirements) may be performed by letting σ in equation 24 be equal to 1,2, ... η, and computing the respective set of formulas, corresponding to equation 26. Useful information concerning the evaluation of coefficients from the continued fraction expansion, equation 8, that would be required for designs based on the computation of the respective open- or short-circuit impedances is given in the Appendix.

Fig. 12. Visual display test equipment

The visual display testing circuits, employing standard networks, have been used satisfactorily for many years for various applications, but their use for adjustment and testing of synthesized 4-poles appears to be almost a necessity.

Appendix

Let Z denote any open- or short-circuit impedance of the 4-pole reactance. By dividing the impedance function by $p = j\Omega$, a new function is obtained

$$F = \frac{Z(p)}{p} = F(p^2)$$

Replacing p^2 with a new variable

$$s = -\frac{1}{p^2}$$

produces

$$F_1 = F_1\left(-\frac{1}{s}\right) = F_1(s) = \frac{C_1(s)}{D_1(s)}$$

The ladder-network coefficients, corresponding to Fig. 2, are now available through the use of the following formulas:

$$a_1 = F_1(s_1) = \frac{C_1(s)}{D_1(s)}$$

$$a_2 = -\frac{D_1(s_1)}{C_2(s_1)}$$

$$a_3 = \frac{C_2(s_2)}{D_2(s_2)}$$

$$a_4 = -\frac{D_2(s_2)}{C_3(s_2)}$$

$$C_2(s) = \frac{C_1(s) - a_1 D_1(s)}{s - s_1}$$

$$D_2(s) = \frac{D_1(s) + a_2 C_2(s)}{s - s_1}$$

$$C_3(s) = \frac{C_2(s) - a_3 D_2(s)}{s - s_2}$$

The evaluation of the coefficients continues until the last one, corresponding to the impedance $Z(p)$, is available. The respective functions C and D required for its evaluation, then are constants.

References

1. EIN INTERPOLATIONSPROBLEM MIT FUNKTIONEN MIT POSITIVEN REALTEIL, W. Cauer. *Mathematische Zeitschrift*, Berlin, Germany, vol. 38, 1933, pp. 1–44.

2. THEORIE DER LINEAREN WECHSELSTROMSCHALTUNGEN, Vol. I (book), W. Cauer. Leipzig, Germany, 1941.

3. WELLENFILTER, INSBESONDERE SYMMETRISCHE UND ANTIMETRISCHE, MIT VORGESCHRIEBENEM BETRIEBSVERHALTEN, H. Piloty. *Telegraphen- Fernsprech- Funk- und Fernseh-Technik*, Berlin, Germany, vol. 28, 1939, pp. 363–75.

4. SYNTHESIS OF REACTANCE 4-POLES WHICH PRODUCE PRESCRIBED INSERTION LOSS CHARACTERISTICS, S. Darlington. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 18, 1939, pp. 257–353.

5. NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN (book), H. W. Bode. D. Van Nostrand Company, Inc., Princeton, N. J., 1945.

6. DESIGN OF ELECTRIC WAVE FILTERS WITH THE AID OF THE ELECTROLYTIC TANK, A. R. Boothroyd. *Proceedings*, Institution of Electrical Engineers, London, England, pt. IV, Oct. 1951.

7. TABLES OF FUNCTIONS WITH FORMULAE AND CURVES (book), E. Janke, F. Emde. Dover Publications, Inc., New York, N. Y., 1943.

8. ELECTRICAL ENGINEERS' HANDBOOK (book), H. Pender, K. McIlwain, editors. "Communication—Electronics," John Wiley & Sons, Inc., New York, N. Y., pp. 6–57, 58.

Proposed Standards for Core Test Methods for Toroidal Magnetic Amplifier Cores

AIEE COMMITTEE REPORT

ONE OF THE major problems in the production of magnetic amplifiers is reproducibility. It is essential that reactor cores of magnetic amplifiers be tested and selected prior to winding if good quality control is to be maintained. There have been many approaches to the problem and each company has its own test circuit. Realizing that it would be advantageous to both the core vendors and the users to standardize on core test methods, the Magnetic Amplifier Materials Subcommittee formed the Working Group on Core Matching and Grading to establish standard test circuits. This paper constitutes a summary of the Working Group's activities over a 2-year period and sets forth recommendations for two standard core test circuits for toroidal cores.

To survey the field, questionnaires were sent to 46 companies who are active in the magnetic-amplifier field. The information received from the 18 companies that answered formed the pool of informaton from which the standard test circuits were determined.

The information was evaluated in the light of the following:

1. Reproducibility of results.
2. Correlation of test data with amplifier performance.
3. Simplicity of circuitry.
4. The number of companies employing each of the submitted test circuits.

All of the proposed test circuits, including the circuits presented in this paper, excite the core in such a manner that the resulting flux excursions as a function of time differ from the excursions that take place in actual magnetic-amplifier operation. From this point of view the circuits are not ideal. The departure, however, is acceptable in view of the fact that correlation of test results with amplifier performance is acceptable.

Of the seven types of test circuits proposed, two were selected. They are the constant-current-flux-reset (CCFR) test and the sine-current-excitation (SCE) test. It was impossible to eliminate either one in preference to the other since both are widely used and neither presents outstanding advantages over the other. The operation of both of these circuits has been well described in previous literature.[1–3] The purpose of this paper is to establish conditions of measurement so that test results may be reproduced from one test installation to the next.

The SCE test measures magnetic parameters while the flux is traversing a major dynamic hysteresis loop. The CCFR test, on the other hand, simulates the operation of a single core half-wave magnetic amplifier. Measurements are made while the flux traverses minor dynamic hysteresis loops.

In both test methods magnetic properties are obtained by measuring electrical quantities. The test equipment is calibrated in magnetic units, since core properties are the quantities of interest.

Since the two test methods excite the core in different ways, one should not expect to obtain the same quantitative results. Only in the measuring of peak induction (B_m) and peak induction minus residual induction ($B_m - B_r$) are the test conditions identical in the two circuits.

CCFR Core Test Method

A1.0. GENERAL DESCRIPTION

A1.1. The CCFR core test method provides a means for measurement of core magnetic properties under operative conditions which approximate conditions in several different types of self-saturating magnetic-amplifier circuits. The test employs an excitation current consisting of half-wave sine current pulse of sufficient magnitude to drive the core flux into positive saturation. A d-c demagnetizing force of adjustable magnitude is applied to the core so as to reset flux away from positive saturation during the intervals between pulses of excitation current. The resultant cyclic flux change is measured by means of a sensitive flux volt-

---

Paper 58-71, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 31, 1957; made available for printing November 7, 1957.

The personnel of the Working Group of the Core Matching and Grading Materials Subcommittee of the AIEE Magnetic Amplifier Committee are: V. J. Louden, *chairman*, H. R. Brownell, J. R. Jacquet, E. W. Lehtonen, J. E. Mitch, R. W. Olmstead, R. W. Roberts, W. Schlosser, W. R. Seegmiller, R. D. Teasdale, C. B. Wakeman, C. E. Ward, and G. L. Thompson.

Vol.
Pt. 1
195

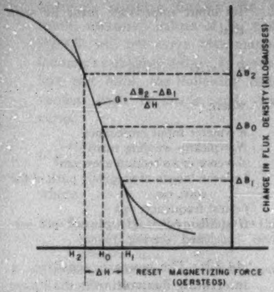


Fig. 1. Test points for the CCFR core test method

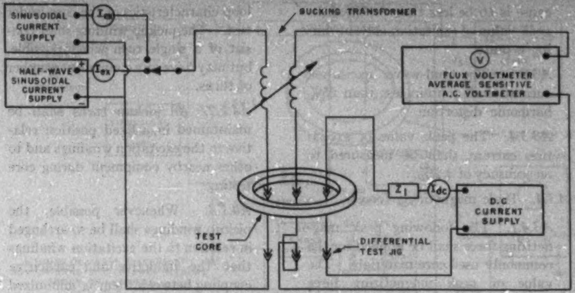

SINUSOIDAL CURRENT SUPPLY
I_ex
HALF-WAVE SINUSOIDAL CURRENT SUPPLY
+
−
I_ex
BUCKING TRANSFORMER
V
FLUX VOLTMETER AVERAGE SENSITIVE A.C. VOLTMETER
Z₁
I_dc
D.C CURRENT SUPPLY
TEST CORE
DIFFERENTIAL TEST JIG

Fig. 2. Circuit for CCFR core test method

meter connected to a separate pickup winding on the core. A low number of excitation, control, and pickup turns are used to facilitate rapid jigging of cores.

A1.2. Test data are obtained by measuring values of induced pickup voltage as a function of the applied demagnetizing control ampere-turns under conditions of half-wave sinusoidal excitation current of constant magnitude. For some purposes these data may best be presented in the form of a curve. For general production, core grading and matching data are obtained only at particular standardized test points. These standard test points are selected as shown in Fig. 1 to determine the bias point H_0 and the gain G in the linear region of the curve. Also, the difference between peak induction and residual induction ($B_m - B_r$) is measured using zero control ampere-turns.

A1.3. The measurement of peak flux density is made by employing full-wave SCE. This measurement can best be made by switching to a full-wave sine current supply instead of a half-wave supply, and using zero control ampere-turns.

A2.0. TEST CIRCUIT

A2.1. A simplified schematic diagram of the test circuit is shown in Fig. 2. The circuit consists of five major blocks: full-wave sine current supply, half-wave sine current supply, d-c control current supply, test jig, and flux voltmeter. The requirements on these various units are given in the following paragraph.

A2.2. A sinusoidal and half-wave sinusoidal current supply shall be provided. These may be separate current supplies, or may be obtained from a common source by suitable switching and rectifying means.

A2.3. Plug-in type test jig:

A2.3.1. A single or double (differential) plug-in type of test jig shall be used. A single test jig shall have a plug-in connector having an excitation winding, a pickup winding, and a control winding.

A2.3.2. If a double or differential type of test jig is employed, the excitation windings of the two plug-in connectors shall have equal numbers of turns and shall be connected in series. The pickup windings of the two plug-in connectors shall have equal numbers of turns, and shall be connected in series opposition. Also, the control windings shall have equal turns and shall be connected in series with the same polarity connections as the excitation windings.

A2.3.3. Provision for cancellation of voltage pickup due to air induction linking the excitation winding with the pickup winding shall be made, if necessary.

A2.4. Flux voltmeter: Either an averaging instrument sensitive to total flux change or an integrating flux voltmeter shall be used to meter the core output voltage.

A3.0. EXCITATION AND CONTROL REQUIREMENTS

A3.1. Voltage source:

A3.1.1. The voltage of the excitation source shall be sinusoidal, with a harmonic content of less than 10% and shall have an rms amplitude greater than 100 times the half-cycle average value of voltage induced in the excitation winding.

A3.1.2. The frequency of the voltage source shall be known within 1%, and if not known within these limits, it shall be measured to an accuracy of 1%.

A3.1.3. Standard test frequencies of 60, 400 or 1,600 cps (cycles per second) may be used within the frequency limitations of the test circuitry and instrumentation. The tolerance on any standard test frequency used shall be ±5%.

A3.2. Excitation windings:

A3.2.1. The excitation winding on the test jig shall consist of a single turn, wherever possible, but shall not exceed six turns.

A3.2.2 All excitation windings shall be kept in a fixed position in relation to the pickup windings, and to nearby circuits and equipment, during testing.

A3.2.3. When differential core matching is desired, two identical excitation windings shall be used, connected in series. Geometric symmetry of the excitation windings in relation to each other shall be maintained.

A3.3. Excitation current:

A3.3.1. The half-wave excitation current shall be essentially sinusoidal. For values of current from 5% to 25% of peak excitation current, the time interval between an instantaneous value of current on the descending portion of the current waveform to an equal value of current on the ascending portion of the next half-cycle of current shall be within ±3% of the corresponding time interval for a perfect half-wave sinusoid.

A3.3.2. During the interval between pulses when the excitation current is nominally zero, the maximum instantaneous current occurring because of rectifier leakage or other cause is to be less than 0.1% of the peak value of excitation current during a pulse.

A3.3.3. The full-wave excitation current shall contain less than 25% harmonic distortion.

A3.3.4. The peak value of excitation current shall be measured to an accuracy of ±2%.

A3.4. Peak magnetizing force:

A3.4.1. The following peak magnetizing force shall be applied for the commonly used core materials. The value of peak magnetizing force applies to materials whose tape thickness is 6 mils or less, when tested at 400 cps or below; or whose tape thickness is from 6 to 14 mils when tested at 60 cps.

Material	Peak Magnetizing Force When Measuring H_0, H_1, H_2, Oersteds	Peak Magnetizing Force When Measuring B_m and (B_m-B_r), Oersteds
3% silicon iron (oriented)	6.0	3.0
50% nickel-iron (oriented)	2.0	1.0
79% nickel-iron	1.0	0.5
Supermalloy	0.5	0.25

where

$H_m=0.4\pi N_e I_m/l$ oersteds
H_m = peak magnetizing force
N_e = number of excitation winding turns
I_m = peak value of excitation current
l = mean core length, cm (centimeters)

A3.4.2. When conditions other than those covered in paragraph *A3.4.1* are employed, the nearest larger value of peak magnetizing force shall be selected, rounded off to one significant figure, i.e., twice the value of the highest sine current coercive force to be expected.

A3.5. D-c control current:

A3.5.1. The rms ripple of the control current shall not exceed 0.25%.

A3.5.2. Control current corresponding to H_0, H_1, and H_2 (Fig. 1) shall be measured to within ±1%.

A3.5.3. Changes in control current corresponding to ΔH(Fig. 1) shall be measured to ±2%.

A4.0. MEASUREMENT AND REQUIREMENTS

A4.1. Pickup windings:

A4.1.1. The pickup winding on the test jig provides an induced voltage for measuring the dynamic hysteresis loop characteristics of the core under test. The pickup winding shall consist of a single turn when possible, but may have any convenient number of turns.

A4.1.2. All pickup turns shall be maintained in a fixed position relative to the excitation windings and to other nearby equipment during core testing.

A4.1.3. Whenever possible, the pickup windings shall be so arranged in relation to the excitation windings that the inductive and capacitive coupling between them is minimized so that, with no core in the jig, no induced voltage appears on the output of the pickup windings with maximum current flowingat the test frequency in the excitation winding.

A4.1.4. If the average induced voltage from the pickup winding under the conditions of paragraph *A4.1.3*, due to mutual coupling between the excitation and the pickup windings, exceeds 0.5% of the average output voltage corresponding to the peak magnetizing force, this mutual inductive coupling between the excitation winding and the pickup winding shall be adjusted so as to reduce the induced voltage to less than 0.5%. If necessary, an air core bucking transformer shall be used for this purpose.

A4.1.5. Where differential core matching is desired, two identical pickup windings shall be used, connected in series opposition. Geometric symmetry of the pickup windings in relation to each other, and to the excitation windings, shall be maintained.

A4.1.6. The average output voltage of the differential jig of paragraph *A4.1.5*, with no cores in the windings and with maximum excitation current flowing at the test frequency, shall be less than 0.5% of the average voltage obtained with only one core in the differential test windings. The mutual inductive coupling between one of the excitation and pickup windings of the differential test jig shall be adjusted to reduce the induced voltage to less than 0.5%.

A4.2. Flux voltmeter input impedance:

A4.2.1. The average sensitive or integrating flux voltmeter shall have negligible loading effect on the core. Its input impedance must be such that

$$R>2\pi f N_s \frac{2A}{l} G\times 10^{-8}$$

where

R = meter input impedance
N_s = pickup winding turns
A = core cross-section area, cm^2
l = mean length of magnetic path of the core, cm
f = test frequency, cps
$G=(\Delta B_2-\Delta B_1)/\Delta H$, gausses per oersted (see Fig. 1)

A4.2.2. The average sensitive or integrating flux voltmeter shall have an accuracy of ±1% of reading.

A4.3. Standard test points:

A4.3.1. Data shall be taken at the following test points:

a. B_m at H_m specified in paragraph *A3.4.1*.

b. B_m-B_r, at control $H=0$.

c H_0 at specified ΔB_0 (paragraph *A4.3.2*).

d. H_1 at specified ΔB_1 (paragraph *A4.3.2*).

e. H_2 at specified ΔB_2 (paragraph *A4.3.2*).

A4.3.2. ΔB_1, ΔB_0, and ΔB_2 as independent variables should be adjusted as follows:

Material	ΔB, Kilogausses		
	ΔB_1	ΔB_0	ΔB_2
3% silicon-iron, oriented	10	15	20
50% nickel-iron	10	15	20
79% nickel-iron	5	7.5	10
Supermalloy	5	7.5	10

The value of ΔB_0 was selected to be approximately one half the maximum flux density swing encountered in usual magnetic amplifier designs. Also, ΔB_1 approximates one third of the design ΔB; and ΔB_2 approximates two thirds of the design ΔB.

A4.3.3. In calibrating the flux voltmeter, which measures total flux change, to read flux density (in gausses) the manufacturer's published gross cross section area of the cores being tested and the following stacking factors shall be employed.

Tape Thickness, Mils	Stacking Factor
1/2	0.50
1	0.75
2	0.85
4 through 10	0.90
11 through 14	0.95

These stacking factors do not necessarily correspond to the manufacturer's stacking factors. As a result, measured values of B_m and $B_m - B_r$ will not agree with manufacturer's published data. To correlate test data, the following conversion must be made, and vice versa:

$$B' = \frac{BAk}{A'}$$

where

B' = flux density corresponding to the manufacturer's published core effective area
A' = manufacturer's published core effective area
k = stacking factor listed in the foregoing
A = published gross core cross-section area
B = flux density read from tester

The same stacking factor is employed in testing equivalent cores from various manufacturers in order that the cores may be tested with the same flux swings, and that the cores may be specified with one description. The tester actually measures flux change, which is the quantity of interest.

A4.4. Measurement of B_m: The peak induction (B_m) shall be measured at the peak magnetizing force specified in paragraph *A3.4.1*, employing full-wave sinusoidal excitation current and zero control current.

A4.5. Measurement of H_0: H_0 shall be measured in oersteds at the value of B_0 specified in paragraph *A4.3.2*, utilizing half-wave sinusoidal excitation current as specified in paragraph *A3.4.1*.

A4.6. Measurement of G in gausses per oersted is computed as follows:

$$G = \frac{\Delta B_2 - \Delta B_1}{H_2 - H_1} = \frac{\Delta B_2 - \Delta B_1}{\Delta H}$$

where ΔB_2 and ΔB_1 are specified in paragraph *A4.3.2* and ΔH is measured, utilizing half-wave sinusoidal excitation current as specified in paragraph *A3.4.1*.

A4.7. Measurement of $B_m - B_r$: $B_m - B_1$ should be measured utilizing zero control current and half-wave sinusoidal current excitation as specified in paragraph *A3.4.1*. *This measurement shall be made directly following the measurement of* B_m *in order to avoid errors resulting from magnetic history.*

A5.0. Differential Core Matching

In matching cores by the differential method, the tolerance of match shall be determined by measuring the differential output voltage between each core and a reference core at values of magnetizing force specified by the user. The difference so measured, as compared with the nominal peak induction of the core shall be a measure of the per-cent tolerance of match between the cores.

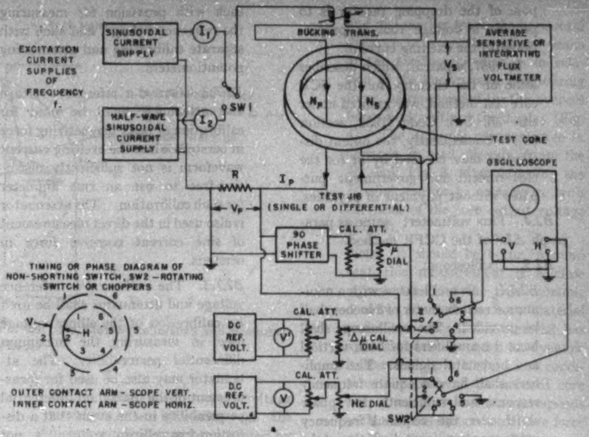

SINUSOIDAL CURRENT SUPPLY
EXCITATION CURRENT SUPPLIES OF FREQUENCY f.
HALF-WAVE SINUSOIDAL CURRENT SUPPLY
I1
I2
SW1
BUCKING TRANS.
Np
Ns
AVERAGE SENSITIVE OR INTEGRATING FLUX VOLTMETER
Vs
TEST CORE
OSCILLOSCOPE
R
Ip
Vp
TEST JIG (SINGLE OR DIFFERENTIAL)
90 PHASE SHIFTER
CAL. ATT.
μ DIAL
V
H
TIMING OR PHASE DIAGRAM OF NON-SHORTING SWITCH, SW2 — ROTATING SWITCH OR CHOPPERS
1
2
3
4
5
6
DC REF. VOLT.
V'
Δμ CAL. DIAL
Hc DIAL
SW2
OUTER CONTACT ARM - SCOPE VERT.
INNER CONTACT ARM - SCOPE HORIZ.

Fig. 3. Basic circuit for SCE core tester

SCE Core Test Method

B1.0. General Description

B1.1. The SCE test method provides for the measurement of four basic parameters of the saturated major dynamic sine current hysteresis loop of a core material for magnetic-amplifier applications. The core parameters measured are generally considered to be significant in determining the performance of the core material in the magnetic-amplifier type of application. These parameters are the a-c peak induction (corresponding to the a-c peak magnetizing force), the squareness ratio (ratio of residual induction to peak induction), the sine current coercive force, and the maximum differential permeability (maximum slope of the sides of the dynamic hysteresis loop).

B1.2. The test method described employs a relatively high impedance excitation source of sine voltage and known frequency, driving the core material under test into relative saturation while maintaining a sinusoidal exciting current. To facilitate rapid jigging of test cores of widely varying sizes, very few excitation and pickup turns are used. Instrumentation is provided for determining the peak magnetizing force, the peak induction, and for observing the output voltage versus exciting current waveform of the core material. Calibration means are provided whereby the sine current coercive force and the maximum differential permeability may be measured from the output voltage versus exciting current waveform. Provision is made for excitation of the core with a half-wave of rectified sine current to facilitate the measurement of flux density change (ΔB or $B_m - B_r$), from which the ratio of residual induction to peak induction may be calculated. Provision is also made for matching of cores on a differential basis.

B2.0. Test Circuit

B2.1. A test circuit such as that shown in Fig. 3 shall be used. The basic circuit consists of the components listed in the following.

B2.2. Same as paragraph *A2.2* of the CCFR method.

B2.3. Plug-in type test jig:

B2.3.1. Same as paragraph *A2.3.1* of the CCFR method.

B2.3.2. If a double or differential type of test jig is employed, the excitation windings of the two plug-in connectors shall have equal numbers of turns and shall be connected in series. The pickup windings of the two plug-in connectors shall have equal numbers of turns, and shall be connected in series opposition.

B2.3.3. Same as paragraph *A2.3.3* of the CCFR method.

B2.3.4. Provision for a noninductive voltage dropping resistor shall be made in series with the source and the excitation windings. The purpose of the dropping resistor is to provide a voltage that is proportional to the exciting current.

In 1957, a patent, which describes some of the circuitry for the SCE core test method, was granted under title 35, U.S. Code (1952, section 66).[4] Consequently, the patent disclosures may be used by or for the Government for government purposes without payment of royalties.

B2.4. Flux voltmeter: same as paragraph *A2.4* of the CCFR method.

B2.5. Oscilloscope:

B2.5.1. An oscilloscope with a minimum screen diameter of 5 inches shall be provided. The oscilloscope shall have d-c amplifiers for both vertical and horizontal signals. The amplifiers shall have adequate frequency response and sufficiently low phase shift[4] over the required frequency range so as to prevent distortion of the observed waveform in excess of one 2/100 of the screen diameter of the cathode-ray tube of the oscilloscope.

B2.5.2. If the output voltage versus exciting current loop characteristics are to be measured by means of scales on the face of the scope rather than by calibrated marker traces, then the linearity of the vertical and horizontal amplifiers and deflection systems shall be within one line width on the face of the oscilloscope.

B2.6. Ninety-degree phase shifter and calibrating attenuators:

B2.6.1. A source of sinusoidal voltage with less than 1% distortion shall be provided for the measurement of maximum differential permeability. This sinusoidal voltage, known as the ellipse voltage, shall be 90 degrees out of phase with the exciting current, and may be derived from the voltage across the excitation circuit dropping resistor by shifting its phase by 90 degrees. The function of the ellipse voltage is to provide a locus of constant differential permeability.

B2.6.2 The ellipse voltage shall be adjustable by separate calibrating and attenuating potentiometers to provide for calibration and measurement of maximum differential permeability readings in terms of gausses per oersted.

B2.7. D-c reference voltage sources:

B2.7.1. Two separate d-c voltage reference sources shall be provided, each with provision for measuring the reference voltage, and each with separate calibrating and attenuating potentiometers.

B2.7.2. One d-c reference voltage and attenuator shall be used for calibration of peak magnetizing force in oersteds when the exciting current waveform is not sufficiently distortion-free to use an rms ammeter for such calibration. The attenuator is also used in the direct measurement of sine current coercive force in oersteds.

B2.7.3. The other d-c reference voltage and attenuator shall be used for calibration of the ellipse voltage used in measuring the maximum differential permeability. The attenuator may also be used for measurement of the maximum differential permeability in the event that a distortion-free ellipse voltage is not obtainable under all test conditions.

B2.8. Rotating switch or phased synchronous choppers:

B2.8.1. A rotating double-pole multiposition switch shall be provided for displaying on the oscilloscope various sequences of the core output voltage versus exciting current, ellipse voltage versus exciting current, d-c reference voltage markers, and zero reference voltage markers.

B2.8.2. In lieu of a suitable rotating switch, two or more single-pole double-throw synchronous choppers may be used, properly phased, to apply the proper sequence of signals to the oscilloscope.

B2.8.3. The frequency of rotation of the switch, or driving frequency of the choppers, shall be not greater than one sixth of the fundamental core test frequency.

B3.0. EXCITATION REQUIREMENTS

B3.1. Voltage source:

B3.1.1. The voltage of the excitation source shall be sinusoidal, with a harmonic content of less than 10%.

B3.1.2. Same as paragraph *A3.1.2* of the CCFR method.

B3.1.3. Same as paragraph *A3.1.3* of the CCFR method.

B3.2. Excitation windings:

B3.2.1. Same as paragraph *A3.2.1* of the CCFR method.

B3.2.2. Same as paragraph *A3.2.2* of the CCFR method.

B3.2.3. Same as paragraph *A3.2.3* of the CCFR method.

B3.3. Exciting current:

B3.3.1. The exciting current shall be sinusoidal with a harmonic content of less than 5% at the value of current corresponding to the peak magnetizing force to be used for the core under test. Series resistance may be used to increase the impedance of the excitation source sufficiently to limit harmonics in the exciting current waveform for the smaller core sizes. When necessary, a series resonant filter shall be used to limit the maximum harmonic content of the exciting current to 5%.

B3.3.2. Suitable means for measuring the peak exciting current with an accuracy of 2% shall be provided. When the harmonic content of the exciting current is sufficiently low, a dynamometer type rms ammeter of at least 0.5% accuracy is recommended.

B3.3.3. Suitable means shall be provided for varying the exciting current to obtain the appropriate peak magnetizing force for the range of core sizes normally encountered. A variable autotransformer is usually satisfactory for this purpose.

B3.4. Peak magnetizing force:

B3.4.1. The following peak magnetizing force shall be applied for the commonly used core materials. The value of peak magnetizing force applies to materials whose tape thickness is 6 mils or less, when tested at 400 cps or less, or whose tape thickness is from 6 to 14 mils or less, when tested at 60 cps.

Materials	Peak Magnetizing Force, Oersteds
3% silicon-iron (oriented)	3.0
50% nickel-iron (oriented)	1.0
79% nickel-iron	0.5
Supermalloy	0.25

B3.4.2. Same as paragraph *A3.4.2* of the CCFR method.

B3.4.3. When a dynamometer type rms ammeter is used to measure the exciting current, the peak magnetizing force shall be calculated from the equation

$$H_m = \frac{0.4\pi N_e I_m}{l} \text{ oersteds}$$

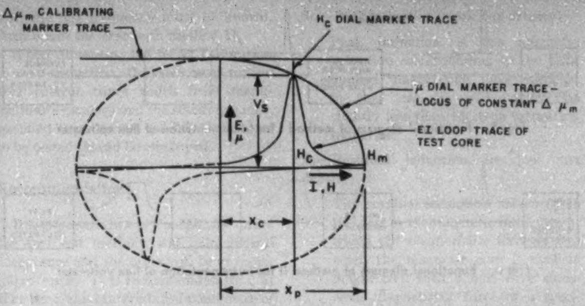

$\Delta\mu_m$ CALIBRATING MARKER TRACE
H_c DIAL MARKER TRACE
μ DIAL MARKER TRACE—LOCUS OF CONSTANT $\Delta\ \mu_m$
EI LOOP TRACE OF TEST CORE
V_S
E, μ
H_C
H_m
I, H
X_C
X_P

Fig. 4. Oscilloscope presentation of E I loop of test core, showing calibrating and measuring marker traces

where

N_e = number of excitation winding turns
I_m = peak exciting current, amperes = 1.414 I_{rms} for a pure sine wave
l = mean core length, cm

B3.4.4. When harmonics in the exciting current waveform are large enough to prevent the measurement of peak magnetizing force with the required accuracy, it may be measured from the oscilloscope display using a d-c reference voltage and calibrated attenuator, using the equation

$$H_p = \frac{0.4\pi N_p}{l} \times \frac{kV}{R} \text{ oersteds}$$

where

N_p = number of excitation winding turns
l = mean core length, cm
k = constant of attenuation of d-c reference voltage source
V = d-c reference voltage, volts
R = excitation circuit dropping resistor, ohms

B4.0. MEASUREMENT REQUIREMENTS

B4.1. Pickup windings:

B4.1.1. Same as paragraph *A4.1.1* of the CCFR method.

B4.1.2. Same as paragraph *A4.1.2* of the CCFR method.

B4.1.3. Same as paragraph *A4.1.3* of the CCFR method.

B4.1.4. Same as paragraph *A4.1.4* of the CCFR method.

B4.1.5. Same as paragraph *A4.1.5* of the CCFR method.

B4.1.6. Same as paragraph *A4.1.6* of the CCFR method.

B4.2. Core output measurements:

B4.2.1. The core output voltage waveform, as appearing on the pickup winding, may be measured by a number of methods to determine the dynamic hysteresis loop characteristics of the core under test.

B4.2.2. All instruments used in making output voltage waveform measurements shall have sufficiently high input impedance so as not to load the core output voltage. The minimum input impedance of this instrument shall be determined as

$$R > 2\pi f N_s \frac{2_A}{l} \Delta\mu_m \times 10^{-8}$$

where

N_s = pickup winding turns
A = core cross section area, cm^2
l = mean core length, cm
$\Delta\mu_m$ = maximum differential permeability of test core in gausses per oersted
f = test frequency, cps

B4.3. Measurement of peak induction or peak flux B_m:

B4.3.1. The measurement of peak induction shall be made by measuring the core output voltage with an average sensitive or integrating flux voltmeter (Fig. 3), using the peak magnetizing force specified in section *3.4.1.* The reading shall be calibrated directly in gausses, based on the core cross section area at a specific test frequency.

B4.4. Measurement of sine current coercive force H_c:

B4.4.1. Sine current coercive force measurements with this test method shall be made by displaying the core output voltage versus exciting current on an oscilloscope and measuring, by means of a calibrated potentiometer and dial marker trace, the horizontal distance between the voltage peaks of the oscilloscope display in relation to the known peak magnetizing force, as represented by the over-all horizontal width of the display (Fig. 4). The dynamic coercive force in oersteds is read directly from the H_c dial, Fig. 3, after adjusting the H_c dial marker to coincidence with the peak of the core output voltage (Fig. 4).

Note: It should be borne in mind that this measurement is a close representation of the true dynamic coercive force of the core material under test only if the material has approximately a rectangular dynamic hysteresis loop. Under these conditions the peak of the induced core output voltage waveform corresponds with the true dynamic coercive force within a few per cent. When nonrectangular dynamic hysteresis loop materials are being measured, however, the peak of the output voltage wave may depart from the true coercive force value by a considerable amount, and the value measured is the H at $\Delta\mu_m$. This value will generally be somewhat different than the true H_c value, but may be used and specified in the same manner as the H_c value as long as the test conditions and the test results obtained are clearly defined.

B4.5. Measurements of maximum differential permeability ($\Delta\mu_m$):

B4.5.1. The measurement of maximum differential permeability shall be made by displaying the core output voltage versus exciting current on an oscilloscope and measuring, by means of a calibrated potentiometer and dial marker trace, the peak value of the output voltage. The maximum differential permeability in gausses per oersteds may be read directly from the μ dial (Fig. 3) after adjusting the μ dial marker to coincidence with the peak of the core output voltage (Fig. 4).

B4.5.2. Calibration of maximum differential permeability shall be accomplished using a d-c reference voltage and calibrated attenuator. When the $\Delta\mu_m$ calibrating marker trace is adjusted to coincidence with the peak value of the ellipse voltage, the maximum differential permeability in gausses per oersted may be calibrated using the equation

$$\Delta\mu_m = \frac{k'V'}{N_s A} \times \frac{l}{0.4\pi N_e I_m} \times \frac{10^8}{w \cos wt}$$

where

N_e = number of excitation turns
N_2 = number of pickup turns
A = nominal core area, cm^3
l = mean core length, cm
I_m = peak excitation current
  $= 1.414\, I_{rms}$ for a pure sine wave
$w = 2\pi \times$ test frequency
V' = d-c reference voltage, volts
k' = attenuation constant of reference voltage attenuator

by setting $wt = 0$ and substituting the appropriate contents into the equation.

B4.5.3. An alternative calibration of maximum differential permeability may be made by adjusting the $\Delta\mu_m$ calibrating marker to coincidence with the peak of the core output voltage wave, and using the equation of *B4.5.2* by setting $wt = (\sin^{-1} H_c/H_m)$ and substituting the appropriate constants in the equation.

B4.6. Measurement of squareness; ratio of residual to peak: Induction (ΔB or B_r/B_m).

B4.6.1. The measurement of squareness shall be made by measuring the change in induction ($B_m - B_r$) with an average sensitive or integrating flux voltmeter with a half-wave sinusoidal excitation whose peak magnetizing force corresponds to that specified in section *B3.4.1.* This measurement of the change in induction shall be made in terms of gausses. *This measurement shall be made directly following the measurement of B_m in order to avoid errors resulting from magnetic history.*

B4.6.2. When the ratio of residual to peak induction is required, rather than ($B_m - B_r$), it may be calculated as

$$\frac{B_r}{B_m} = \frac{B_m - (B_m - B_r)}{B_m} = 1 - \frac{B_m - B_r}{B_m}$$

with the value of peak induction, B_m, measured as in section *B4.3.1.*

B5.0. DIFFERENTIAL CORE MATCHING

B5.1. When matching of cores by the differential method is desired, the over-all match shall be determined as a percentage, plus or minus, using one of the cores in each matched set as a reference. The tolerance of match shall be determined by measuring the differential output voltage between each core and the reference core with an average sensitive or integrating flux voltmeter; see Fig. 3. The difference in induction so measured, as compared with B_m of the reference core, shall be a measure of the per-cent tolerance of match between the cores.

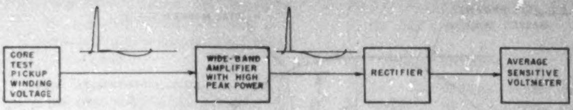

CORE TEST PICKUP WINDING VOLTAGE
WIDE-BAND AMPLIFIER WITH HIGH PEAK POWER
RECTIFIER
AVERAGE SENSITIVE VOLTMETER

Fig. 5. Functional diagram of method I for instrumentation of flux voltmeter

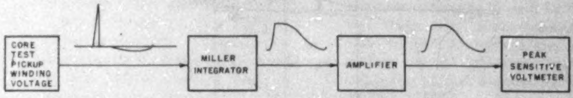


Fig. 6. Functional diagram of method II for instrumentation of flux voltmeter

Flux Voltmeter Instrumentation

The output of the core test circuits described in the foregoing is obtained from a pickup winding on the core test jig. The instantaneous induced voltage in the pickup winding is proportional to the rate of change of flux. This instantaneous induced voltage is fed into a flux voltmeter. The function of the flux voltmeter is to integrate the induced voltage and amplify as necessary to give a meter reading proportional to the flux change.

Two methods of flux voltmeter instrumentation have been used. Both methods are theoretically correct. The difference in the two methods of instrumentation is in the means of accomplishing the integration and the type of amplifier required. The first method,[6] is to amplify the induced voltage, rectify it, and read the output on an average sensitive voltmeter. The second method[7] is to integrate the induced voltage, then amplify it, and read the output on a peak sensitive voltmeter. In method I the integration is accomplished by the averaging action of the meter. In method II the integration is accomplished by an integrating network. A detailed description of each method has been published in the literature.[6,7] Only a short summary of each will be given here.

A functional diagram of method I is shown in Fig. 5. The induced voltage is amplified by a wide-band vacuum-tube amplifier. A typical waveform of the type encountered during core tests is shown in Fig. 5. The ratio of peak to average voltage may be as high as 50 to 1. For accurate measurements the wide-band amplifier must have sufficient peak-handling capabilities to reproduce faithfully these high peak voltages. Therefore, special consideration must be given in selecting this amplifier to be certain that it does not saturate on these high peak inputs. The output of the amplifier is rectified and fed into an average sensitive (d'Arsonval) voltmeter.

A functional diagram of the method II is shown in Fig. 6. The induced voltage is fed into a Miller-type integrating amplifier circuit which is equivalent functionally to a resistance-capacitance voltage integrator and subsequent amplifier. The integrated voltage is amplified by a conventional linear amplifier. It is evident from an examination of the waveforms shown in Fig. 6 that the amplifier requirements are not as stringent for method II as they are for method I, since there are no high peaks in the integrated voltage waveform. The output of the amplifier is fed into a peak-sensitive voltmeter. The peak-sensitive voltmeter may be constructed from a voltage doubler rectifier circuit with a capacitor across the output which charges to the peak-to-peak output. This peak output can then be metered with an average sensitive (d'Arsonval) voltmeter.

Since both meters measure the total flux change in a core during each cycle of the excitation frequency, the meters can best be calibrated by utilizing a core in which the total flux change for a given excitation is known. The excitation must be sufficient to drive the core well into saturation. A Supermalloy core excited with a peak excitation of 10 oersteds is suggested. The total flux change, when excited from minus to plus 10 oersteds may be determined by normal d-c methods.

The calibration core may be placed in the test jig and excited with the magnetizing force for which the flux swing of the core is known. The meter may then be calibrated in terms of the known flux change.

With method I the meter reading is proportional only to average voltage. When converting the reading to units of flux change, both during calibration and test, the test frequency must be known. Such is not the case with method II.

When converting the meter calibration to read induction, the core cross-section area (height times width from manufacturer's catalog) and the stacking factor, specified in paragraph *A4.3.3*, of the cores to be tested should be employed.

Recommendations

It is expected that as time passes, one of the two test methods will gain general acceptance and the other will be of minor importance. It is recommended that, at that time, the Materials Subcommittee of the Magnetic Amplifier Committee review this specification with an aim of having only one standard test circuit. Specified test conditions, such as peak magnetizing force, should be reviewed in the light of making them applicable to only one test circuit rather than to two.

Glossary of Special Terms

The following list defines certain special terms used in this paper. An effort has been made to make these definitions compatible with the American Standards Association definitions of electrical terms,[8] and the proposed standard terms and definitions for magnetic amplifiers of the AIEE Magnetic Amplifier Committee.

H_m—Peak magnetizing force

> Peak magnetizing force is the maximum value of applied magnetomotive force per mean length of path of the core.

B_m—Peak induction (peak flux density)

> Peak induction is the magnetic induction corresponding to the peak applied magnetizing force specified in this test. It will usually be slightly less than the true saturation induction.

B_r—Residual induction (residual flux density)

> The residual induction measured in this test is the magnetic induction at which the magnetizing force is zero when the magnetic core is cyclicly magnetized with a half-wave sinusoidal magnetizing force of a specified peak magnitude. This use of the term residual induction differs from the standard definition which requires symmetrically cyclicly magnetized conditions.

Sine current magnetizing force

> The sine current magnetizing force is the applied magnetomotive force per unit length for a core symmetrically cyclicly magnetized with a sinusoidal current.

$\Delta\mu$—Sine current differential permeability

> The sine current differential permeability is the slope of the sides of the dynamic hysteresis loop obtained with a sine current magnetizing force.

$\Delta\mu_m$—Maximum sine current differential permeability

> The maximum value of sine current differential permeability obtained with a specified sine current magnetizing force.

H_c—Sine current coercive force

> The sine current coercive force is the instantaneous value of sine current magnetizing force at which the dynamic hysteresis loop passes through zero induction.

ΔB—Delta induction (delta flux density)

> Delta induction is the change in induction (flux density) when a core is in a cyclicly magnetized condition.

H_0, H_1, H_2,—Bias points; see Fig. 1

ΔB_0, ΔB_1, ΔB_2,—Independent variables; see Fig. 1

G—Gain; see Fig. 1

References

1. MAGNETIC AMPLIFIER GAPLESS CORE TESTS, J. R. Conrath. *Electronics*, New York, N. Y., Nov. 1952, pp. 119–21.

2. MAGNETIC CHARACTERISTICS PERTINENT TO THE OPERATION OF CORES IN SELF-SATURATING MAGNETIC AMPLIFIERS, R. W. Roberts. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

3. A METHOD OF MEASURING DYNAMIC MAGNETIC PROPERTIES OF CORE MATERIALS, J. D. Patrick. *NAVORD Report 940-1456*, U. S. Naval Ordnance, Washington, D. C., Oct. 6, 1954.

4. A-C PERMEABILITY AND HYSTERESIS ANALYSER, J. D. Patrick, A. Janssens. *U.S. Patent no. 2,805,390*, 1957.

5. DYNAMIC HYSTERESIS LOOP MEASURING EQUIPMENT, H. W. Lord. *AIEE Transactions* (*Electrical Engineering*), vol. 71, pt. I, Sept. 1952, pp. 269–72.

6. DESIGN CRITERIA FOR A PRACTICAL FLUX-RESET CORE TESTER, J. R. Jaquet. "Conference on Magnetic Amplifiers," *AIEE Special Publication T-86*, July 1956, pp. 74–105.

7. A DELTA-B INDICATOR, H. W. Lord. *AIEE CP 56-649*, 1956 (available on request).

8. AMERICAN STANDARD DEFINITIONS OF ELECTRICAL TERMS. *AIEE Standards* (C42 series).

Transistor-Oscillator Induction-Motor Drive

W. H. CARD
NONMEMBER AIEE

Synopsis: Transistor and saturable-core-transformer oscillators driving small 2-phase induction motors provide, in effect, a d-c motor without a commutator. The inductive motor load on the oscillators requires that the transistors carry current in both directions and this is aided by shunting the transistors with rectifiers. Simple interconnection of two oscillators provides the 2-phase system. Operation of an experimental drive in the range from 125 to 500 cycles per second resembles behavior of a separately excited d-c motor except that motor shaft rotation is not reversed by reversing the d-c source. The harmonics in the square-wave motor voltages usually have negligible effect.

ORDINARY d-c machines with commutators have been highly developed but the risk of brush-arcing in explosive atmospheres and rapid brush wear especially at high altitudes makes attractive the idea of a d-c machine without a commutator. Power transistors now available have good switching properties so there is the possibility of performing the switching function of the commutator by means of transistors. However, rather than directly modifying a d-c machine to replace the commutator with transistor circuits, a simpler scheme appears to be to build power oscillators to drive conventional a-c machines. A 2-phase transistor oscillator to drive a 2-phase induction motor is described here.

In the experimental system reported here four power transistors, four recti-

---

Paper 58-898, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted October 14, 1957; made available for printing May 13, 1958.

W. H. CARD is with Syracuse University, Syracuse, N. Y.

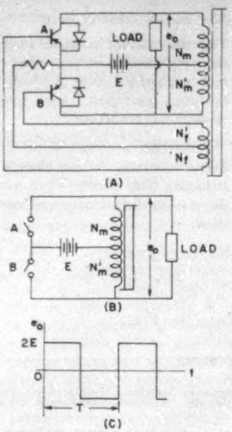


Fig. 1. Transistor oscillator

A—Basic circuit
B—Simplified circuit model
C—Output voltage waveform

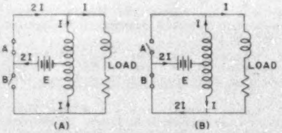


Fig. 2. Distribution of load-component currents

A—At instant before switching
B—At instant after switching

fiers, two saturable transformers, and a single d-c source are used to drive a small 2-phase servomotor. The resulting performance is somewhat like that of a separately excited d-c motor except that rotation is not reversed by reversing the supply voltage. One advantage of the system over conventional 2-phase servomotor drives is that there is no fixed-phase power loss. The important contributions of this paper are believed to be the novel phase-locking connections and the calculated and experimental motor performance curves. Subjects for further investigation are the behavior of other types of a-c motors with the same drive and the generator behavior of the system.

Transistor Power Oscillator

Modified forms of the transistor square-wave oscillator described by Royer[1] are used to drive the motor phases. The circuit shown in Fig. 1(A) differs from Royer's circuit in that rectifiers are connected across the transistors and the load is connected across the transformer main winding (N_m, N_m') rather than to a separate winding. The reasons for this load connection and the phase-locking of two oscillators are explained in the next section. This section reviews the operation of the circuit and discusses the function of the rectifiers when driving inductive loads.

The operation of the oscillator has been explained in terms of the switching behavior of the transistors.[1,2] Provided the base-circuit drive on the transistors from the feedback winding (N_f,N_f') is sufficient to drive the transistors to short circuit, the oscillator may be represented by the circuit model of Fig. 1(B) with the transistors and parallel rectifiers replaced by switches. The output voltage e_0 is a square wave as shown in Fig. 1(C). If a square-loop-core transformer is used and switching occurs when the core saturates, the frequency is given by

$$f=\frac{1}{T}=\frac{E}{4N_m\phi_m} \qquad (1)$$

where ϕ_m is the saturation core flux. Frequency and output voltage are ordinarily little changed for resistive or inductive loads, provided the load current does not approach too closely the limit set by transistor gain and the feedback circuit. The oscillator output is nearly ideal for driving an ordinary a-c motor because, for varying voltage and frequency proportional to voltage, the magnetic saturation in the motor remains nearly constant.

If the oscillator is used to drive an inductive load such as a motor, the load current cannot change magnitude instantaneously, hence a path must be provided in the oscillator circuit to allow the load current to continue to flow when the square-wave oscillator voltage switches polarity. Assuming perfect coupling between N_m and N_m', Fig. 2 shows the distribution of the load components of current for an inductive load at instants before and after switching for the circuit model of Fig. 1(B). Since many available power transistors have less current gain in the inverted than the normal direction, the rectifiers shown in Fig. 1(A) are used to carry at least part of the load current for a short interval after each voltage reversal.

At the instant after switching, the source current is reversed as shown in Fig. 2(B). This means that some of the stored energy in the reactive load is fed back into the source, so the source must be capable of receiving as well as supplying energy. The situation is similar to that of an ordinary a-c reactive load where the instantaneous power flow is negative during part of the cycle.

The transformer windings N_m and N_m' should have good coupling because any leakage reactance results in increased transistor dissipation.[2] It is important to note that increasing transformer leakage reactance increases transistor dissipation, but adding load reactance mainly requires that the "switch" carry current in either direction.

Two-Phase Oscillator

Each motor phase is driven by an oscillator, and the two oscillators are locked in synchronism with their voltages in quadrature. This section describes the phase-locking connection used. Other schemes have been used for phase-locking[3] and it is possible to build oscillators that are intrinsically polyphase,[4] but the method reported here has the virtue of simplicity.

Fig. 3(A) shows a simplified circuit model of the system used. The feedback windings are not shown, the transistors and parallel rectifiers are replaced by the switches A, B, C, and D, and the transformers are assumed to have square-

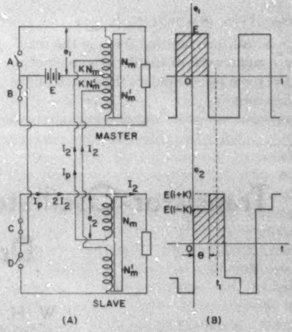


Fig. 3. Two-phase oscillator

A—Simplified circuit model
B—Transformer voltage waveform

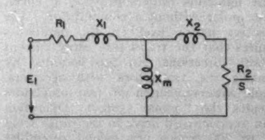


Fig. 4. Induction-motor equivalent circuit

loop cores with negligible loop width. The master oscillator is like the one discussed in the preceding section except for two additional taps at KN_m and KN_m' on the transformer main winding. The slave oscillator is supplied from the same source as the master oscillator and is otherwise the same except that instead of returning the center point on the transformer main winding to the source, the winding is split and returned to the source through the taps on the master oscillator transformer.

Consider first steady-state operation with the two oscillators operating at the same frequency and in phase quadrature. The two oscillator transformer voltage waveforms are shown in Fig. 3(B). The master oscillator waveform is undisturbed by the interconnection, and the (shaded) volt-second area per half-cycle is determined as usual by the saturation volt-seconds of the square-loop-core transformer. In each half-cycle the slave oscillator transformer voltage is first reduced by the factor K, then increased by the factor K of the master oscillator transformer voltage. The (shaded) volt-second area is the same as that of the master oscillator. Thus, if the transformers have the same volt-second area as viewed from N_m, N_m', the requirements are met for both oscillators to operate at the same frequency.

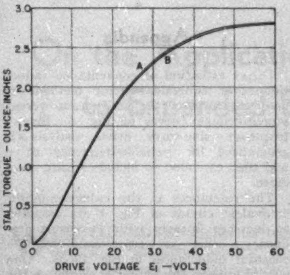


Fig. 5. Calculated stall torque versus drive voltage for 400-cycle servomotor with square-wave drive and frequency proportional to voltage. Curve A neglects and curve B includes effect of harmonics

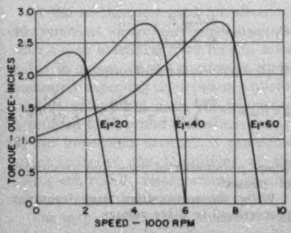


Fig. 6. Calculated speed-torque curves for low-rotor-resistance induction motor with frequency proportional to drive voltage

Provided the two oscillators have the same frequency they will lock in phase quadrature because any change of phase brings compensating action. For example, suppose the slave oscillator tends to speed up. Then in Fig. 3(B), θ becomes greater so that the slave oscillator transformer voltage $E(1-K)$ is applied for more than a quarter period and $E(1+K)$ for less than a quarter period. Since the volt-second area is fixed, the half-period is lengthened and the slave oscillator slows down. In the same way, if the slave oscillator tends to slow down it is speeded up.

Except at the instant of switching the slave oscillator does not appreciably load the master oscillator. Consider time t_1 when switches B and C are closed. The load-component currents I_2 of the slave oscillator are shown in Fig. 3(A). Current I_2 produces no net ampere-turns on the master oscillator transformer.

At the instant before slave oscillator switching, the current in switch C increases by some amount I_p because the slave transformer saturates. This component of current does load the master oscillator and flows through the master oscillator transformer in a direction to increase the required current through switch B. Consequently, there is some tendency for the slave oscillator to cause premature switching of the master oscillator.

The "stepped" waveform of the slave oscillator transformer voltage does not affect the slave oscillator load voltage. It remains square and of the same magnitude as that of the master oscillator. However the stepped waveform appears in the slave oscillator feedback winding, hence the transistor base current is affected. This means that K should be small, for if K approaches unity then during the first quarter cycle of the slave oscillator there would be little base current to hold the transistor (switch C) at short circuit. In the experimental circuit $K=0.25$ is used.

While it may be possible to use the phase-locking principle described here if the loads are connected to separate windings on the two transformers, there are two drawbacks. If the slave oscillator has a separate load winding, net ampere-turns are applied by the slave oscillator to the master oscillator transformer.

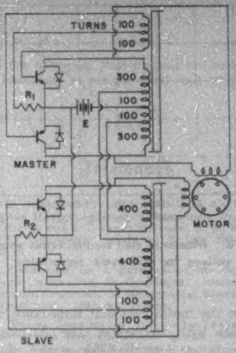


Fig. 7. Experimental 2-phase oscillator induction-motor drive

Transistors—2N57
Rectifiers—IN92
Cores—Magnetics, Inc., 50018-1A

Also, the slave oscillator load voltage has a stepped waveform like Fig. 3(B) instead of the square waveform.

Calculated Motor Performance

Two special problems must be considered when the 2-phase transistor oscillator described in the foregoing is used to drive an induction motor. First, motor performance with square-wave voltage drive is discussed. Second, since the oscillator frequency and output voltage both vary directly with source voltage, performance with this type of drive is reviewed. Calculated curves for low-slip and high-slip motors illustrate the discussion.

The induction-motor equivalent circuit shown in Fig. 4 is used to predict motor behavior. The resistances are assumed constant regardless of speed and frequency, and the reactances are assumed to vary directly with frequency. Torque calculation at any particular drive voltage and frequency is made by conventional methods.[5]

Induction-motor operation with square-wave voltage drive differs from that with sinusoidal drive in two respects. When comparing performance with a sine-wave voltage drive e_1 and a square-wave voltage drive e_2 which has a fundamental component equal to e_1, the power input will be greater and the torque less.

If equal resistances are driven by e_1 and e_2 the ratio of power inputs p_1 and p_2 is

$$p_1/p_2=(2\sqrt{2}/\pi)^2=1.22 \qquad (2$$

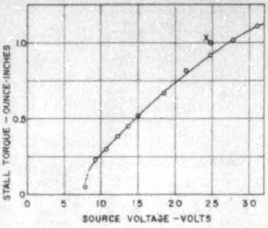


Fig. 8. Measured stall torque versus source voltage for 400-cycle servomotor

Since the motor terminal impedance increases with frequency, the harmonic currents are less than for a resistance. Hence the motor electric power input is never more than 22% greater than for the corresponding sinusoidal drive.

By superposition the total torque is the algebraic sum of the torques contributed by the harmonics. For 2-phase drive with identical but nonsinusoidal voltages in quadrature, the fundamental, fifth, ninth, etc., harmonics contribute positive torques and the third, seventh, eleventh, etc., harmonics contribute negative torques. Analysis of the equivalent circuit (see the Appendix) shows that the third-harmonic stall torque approaches one third, and the sum of the stall torques approaches $\pi/4=0.785$ of the fundamental alone for the worst condition, which is very low frequency.

Calculated performance of a high-slip 2-phase 8-pole induction motor (servomotor) with square-wave drive is shown in Fig. 5. The motor parameters at 400 cycles are $R_1=10$ ohms, $X_1=32$ ohms, $X_m=64$ ohms, $X_2=32$ ohms, $R_2=70$ ohms. The drive frequency f varies as the square-wave voltage amplitude E_1 by $f=10E_1$. In the low-voltage region, the stall torque increases as the third power of drive voltage because of the short-circuiting effect of the magnetizing reactance. The component torques caused by square-wave voltage harmonics decrease the stall torque by a nearly constant amount throughout the range up to rated voltage and frequency. The system is not reversible as is a separately excited d-c motor but, except at low voltage, has stall torque nearly proportional to drive voltage.

Calculated speed-torque curves, neglecting voltage harmonics, for a low-rotor-resistance induction motor are shown in Fig. 6. Except for $R_2=15$ ohms, the same drive and motor parameters are used here as in Fig. 5. The maximum torque and slip speed for maximum torque are nearly constants independent of drive voltage. By increasing the drive voltage from zero, nearly maximum torque is available for starting. At low torque the speed is nearly proportional to drive voltage as in a separately excited d-c motor.

Experimental Results

The experimental oscillator shown in Fig. 7 operates at about 400 cycles per second with a 24-volt source. The collector-to-collector square-wave voltage of about 44 volts amplitude indicates transistor and winding resistance drop of about 2 volts referred to the battery, and a fundamental output voltage of $44\ (2\sqrt{2}/\pi)=40$ volts. Frequency is variable from 125 cycles, or lower on light load, to 500 cycles where maximum-rated collector-to-emitter voltage of 60 volts is reached. The rectifiers effectively clip the voltage spikes that otherwise appear on the output square wave. The transformers used are larger than necessary and improved oscillator design should extend the frequency range.

Neither cores nor transistors are specially matched in the 2-phase oscillator. Because the two oscillators, operated independently, have frequencies about 5% different, they lock in synchronism but not in phase quadrature. However, sufficient adjustment is possible by varying R_1 and R_2 to get good quadrature operation over the operating range.

The 2-phase oscillator was used to drive a small 400-cycle 8-pole servomotor requiring 6.2 watts per phase at 40 volts per phase. The stall-torque versus source-voltage curve shown in Fig. 8 agrees in form with the calculated curve shown in Fig. 5. The point X shows a test for sine-wave drive with voltage equal to the square-wave fundamental indicating the voltage harmonics have a small effect, as predicted.

To verify the theory for a low-rotor-resistance machine, a 60-cycle motor (Diehl *FP-25*) was driven at frequencies up to 500 cycles. (The voltage at 60 cycles was about 3 volts instead of 22 volts rated for the motor.) The speed-torque curves shown in Fig. 9 indicated the general behavior of nearly constant pull-out torque as predicted by the theoretical plot in Fig. 6.

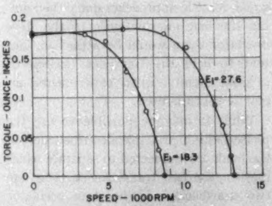


Fig. 9. Measured speed-torque curves for 60-cycle servomotor

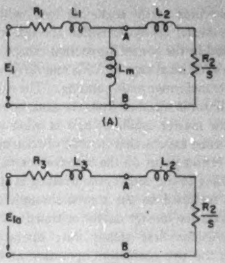


Fig. 10. Modified induction-motor equivalent circuits

Appendix

Direct extension of conventional induction-motor equivalent-circuit methods is used to calculate the stall-torque versus source-voltage curves of Fig. 5. Single-frequency sine-wave drive analysis is generalized for variable-frequency drive and then extended to include square-wave drive.

The reactances in the induction-motor equivalent circuit of Fig. 4 are assumed to increase linearly with increasing frequency, hence are represented by the constant inductances in Fig. 10(A). One way to analyze this circuit is to replace the network to the left of nodes A and B by the Thévenin equivalent circuit shown in Fig. 10(B) by setting

$$E_{1a}=\left|\frac{j\omega L_m}{R_1+j\omega(L_1+L_m)}\right|E_1 \tag{3}$$

and

$$R_3+j\omega L_3=\frac{j\omega L_m(R_1+j\omega L_1)}{R_1+j\omega(L_1+L_m)} \tag{4}$$

With slip $S=1$, the stall torque is

$$T_1(E_1,\omega)=\frac{pq}{2\omega}\left\{\left|\frac{E_{1a}}{R_3+R_2+j\omega(L_3+L_2)}\right|\right\}^2 R_2 \tag{5}$$

where p = number of poles and q = number of phases.

With sine-wave drive and frequency proportional to drive voltage

$$\omega=cE_1 \tag{6}$$

and the curve of stall torque versus supply voltage (Fig. 5 curve A) is plotted by

Vol. 7
Pt. 1
1958

using equations 3 through 6 for a succession of values of E_1.

For a single-frequency square-wave drive of amplitude $\pi E_1/2\sqrt{2}$, the drive voltage may be represented by the Fourier series

$$e_1(t)=\sqrt{2}E_1\left(\cos\omega t-\frac{1}{3}\cos 3\omega t+\frac{1}{5}\cos 5\omega t\ldots\right) \quad (7)$$

Each component contributes torque and these are calculated individually and summed algebraically.

Of special interest is the shape of the stall-torque versus drive-voltage curve for very low voltage. First for sine-wave drive, from equations 3 through 6,

$$E_{1a}\approx\left|\frac{j\omega L_m}{R_1}\right|E_1 \quad (8)$$

$$E_{1a}\approx\frac{cL_m}{R_1}E_1^2 \quad (9)$$

Since

$$R_3+j\omega L_3\ll R_2 \quad (10)$$

the stall torque at low voltage is

$$T_1\approx\frac{pq}{2cE_1}\left(\frac{cL_mE_1^2}{R_1R_2}\right)^2R_2 \quad (11)$$

$$T_1\approx\frac{pqcL_m^2}{2R_1^2R_2}E_1^3 \quad (12)$$

indicating a very slow initial rise of stall torque with increasing supply voltage.

Now consider the extreme case of square-wave drive of such low voltage and frequency that even the harmonics are within the range of approximate validity of equation 9. The nth harmonic of the square wave has rms value E_1/n and the Thévenin equivalent voltage for the nth harmonic is

$$E_{na}\approx\left|\frac{jn\omega L_m}{R_1}\right|\frac{E_1}{n} \quad (13)$$

$$E_{na}\approx E_{1a} \quad (14)$$

Hence from equations 5, 10, and 14 the ratio of nth-harmonic stall torque to fundamental stall torque is

$$\left|\frac{T_n}{T_1}\right|\approx\frac{\frac{pq}{2n\omega}\left(\frac{E_{1a}}{R_2}\right)^2R_2}{\frac{pq}{2\omega}\left(\frac{E_{1a}}{R_2}\right)^2R_2} \quad (15)$$

$$\left|\frac{T_n}{T_1}\right|\approx\frac{1}{n} \quad (16)$$

All the harmonic stall torques are summed algebraically to get the total stall torque

$$T=T_1\left(1-\frac{1}{3}+\frac{1}{5}\ldots\right)=\frac{\pi}{4}T_1$$

References

1. A SWITCHING TRANSISTOR D-C TO A-C CONVERTER HAVING AN OUTPUT FREQUENCY PROPORTIONAL TO THE D-C INPUT VOLTAGE, G. H. Royer. *AIEE Transactions*, vol. 74, pt. I, July, 1955, pp. 322–26.

2. OPERATION OF A SATURABLE-CORE SQUARE WAVE OSCILLATOR, D. C. Moger. *Proceedings*, National Conference on Aeronautical Electronics, Dayton, Ohio, 1956, pp. 481–92.

3. PHASE LOCKING OF SWITCHING-TRANSISTOR CONVERTERS FOR POLYPHASE POWER SUPPLIES, A. G. Milnes. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 587–92.

4. TRANSISTOR CIRCUIT ENGINEERING (book), R. F. Shea, editor. John Wiley & Sons, Inc., New York, N. Y., 1957, pp. 427–29.

5. ELECTRIC MACHINERY (book), A. E. Fitzgerald, C. Kingsley, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., 1952, pp. 392–405.

On the Application of Zener Diodes to Expanded Scale Instruments

A. J. CORSON
MEMBER AIEE

WITH THE ADVENT of closer regulation of industrial electric circuits comes the need for high-accuracy limited-range measurement about a mean value, representative of normal circuit conditions. This applies with particular force to voltage and frequency measurement, and there are many instances in which it is equally true in the measurement of current and power.

Of the mechanisms used for electrical measurement, all are inherently long range with a span extending from zero to some value in excess of the maximum to be measured, expecting the frequency meter. In this exceptional case, the frequency is reduced to a current difference or ratio, by an impedance network, a procedure which removes the limitations of the spring-controlled measurement mechanism in respect to measurement span.

The span of voltmeters, ammeters, and wattmeters may be reduced and the sensitivity increased by prestressing the control spring in the down-scale direction. A voltage (for example, V_1) is required to deflect the movement back to zero, and the voltmeter may then be calibrated over the range V_2-V_1, where V_2 is the voltage required for full-scale deflection. This expedient originates several problems, however, notably, loss of mechanical zero as a check on the condition of the instrument, loss of independence of zero and full-scale adjustments, and loss in restoring-torque gradient. For these reasons, the gain in sensitivity, or readability, is seldom accompanied by a commensurate gain in accuracy. Indeed, unless the instrument is designed for suppressed-zero operation, this method is equivalent to simply magnifying a limited portion of the scale range.

In January 1956, E. E. Helin of the Bell Laboratories suggested the use of the Zener voltage of silicon diodes for expanded scale voltmeters and demonstrated a series d-c circuit in which the reversed diode blocked the instrument current at values below the desired measurement range. This voltmeter circuit was simple and effective, and resulted in a 4-to-1 sensitivity increase, using a standard microammeter calibrated in volts. We then proceeded with a study of the use of Zener diodes in a-c voltage, current, and frequency measurement. The results show that limited-range high-accuracy measurement can be achieved wherever the need for it may exist.

The Silicon Diode

In 1952, Pearson and Sawyer[1] of the Bell Laboratories reported the preparation of a new type of p-n junction silicon diode by alloying acceptor of donor impurities with n- or p-type silicon. The unique features of this diode were: 1. reverse currents as low as 10^{-10} amperes, 2. rectification ratios as high as 10^8 at one volt, 3. a stable Zener characteristic in which $\Delta(\log I)/\Delta(\log V)$ was as high as 1,500 over several decades of current, and 4. ability to operate at ambient temperatures as high as 300 C (degrees centigrade).

In 1954, D. H. Smith,[2] also of the Bell Laboratories, proposed the use of the Zener voltage as a reference standard and

---

Paper 58-934, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 21, 1958; made available for printing May 9, 1958.

A. J. CORSON is with the General Electric Company, West Lynn, Mass.

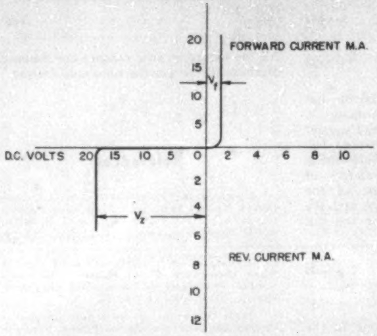


Fig. 1. D-c voltage current characteristic of silicon diode

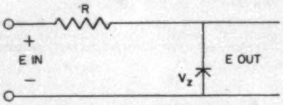


Fig. 2 (left). D-c voltage reference circuit using diode reverse current characteristic

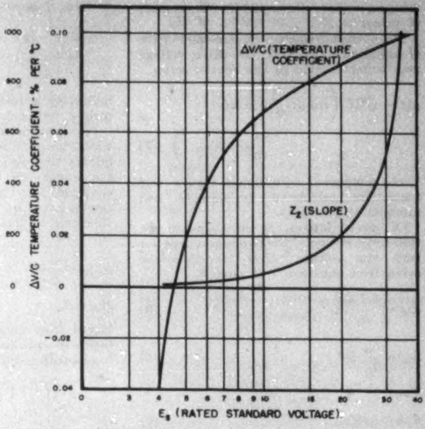


Fig. 3. Impedance and temperature coefficient as function of rates voltage for silicon fused-alloy junction diodes

presented data on temperature coefficient, resistance, and breakdown voltage. The availability of such diodes was subsequently announced by several manufacturers, notably National Semiconductor Products and Transitron Electronic Corporation.

The typical d-c voltage-current characteristic of a silicon diode is shown in Fig. 1. When used for voltage reference, the essential parameters are: 1. The Zener voltage drop V_z, 2. the impedance in both the nonconducting and conducting reverse voltage range, 3. temperature coefficient, and 4. allowable dissipation.

A typical d-c voltage-reference circuit is shown in Fig. 2. The value of the current-limiting resistor R is chosen so that the Zener current $I_z \cong E-V_z/R$ will bias the diode to a satisfactory value as determined primarily by the allowable Zener impedance and secondarily by the desired (exact) Zener voltage. The temperature coefficient depends on the voltage rating as shown in Fig. 3. It is negligible for 4- to 5-volt diodes and is as high as 0.10% per C for 40-volt diodes. Since the characteristic of any one voltage rating is linear, compensation can be used if higher Zener voltages are mandatory. A better but more expensive method of obtaining higher reference voltages is by use of several low-voltage diodes in series. This procedure reduces the Zener impedance as well as the temperature coefficient. The dissipation capacity of all diodes was approximately 150 milliwatts with a recommended derating of one milliwatt per C above a 25 C ambient temperature. Voltage ranges of 5 to 18 volts are available. The stability of these diodes, in relation to the requirements of this application, appears to be excellent. No observable drift has been noted in 10 months' observation of several of them.

When connected to an a-c source through a series resistor, the reverse voltage drop across the diode is the applied voltage until the applied voltage equals the Zener voltage. The Zener voltage is then held until the applied voltage decreases to Zener value, as shown in Fig. 4. At that point the diode again becomes an open circuit, and the voltage drop is coincident with the applied voltage until reversal of the voltage wave drop takes place. (It is assumed in this discussion that the forward current is blocked by another rectifier.) It follows that the average value of Zener voltage is a function of the peak applied voltage, increasing as the applied voltage increases from V_0 and V_0'. In order to establish the limits of satisfactory regulation, the average clipped voltage is derived in terms of the peak applied and Zener voltages as follows. Let

V_z = Zener voltage (see Fig. 5)
V_a = average of clipped voltage
V_0 = peak applied voltage
$e = V_0 \sin$

$$V_a=\frac{1}{\pi}\left[2\int_0^{\alpha} V_0' \sin \alpha d_\alpha + V_z(\pi-2\alpha')\right]$$

which after integration and reduction results in:

$$V_a=\frac{V_0[2(1\sqrt{1-K^2})+K(\pi-2\sin^{-1}K)]}{\pi}$$

where

$K=V_z/V_0$

The plot of this relationship (Fig. 5) shows that this method of stabilization becomes more effective as the ratio of peak applied to Zener voltage is increased. Exact stabilization is unnecessary in such applications as voltage and frequency measurement, nevertheless, we found it desirable to maintain V_0/V_z ratios of 5 or better to get effective results.

Circuit for A-C Expanded Scale Voltmeters

After a study of the possibilities, it was decided to evaluate the diode circuit shown in Fig. 6(A) for a-c voltage meas-

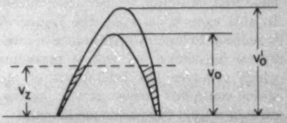


Fig. 4. Relationship between applied sinusoidal voltage and the regulated voltage. Regulation is inexact because of addition of crosshatched areas as applied voltage is increased from E_{o1} to E_{o2}

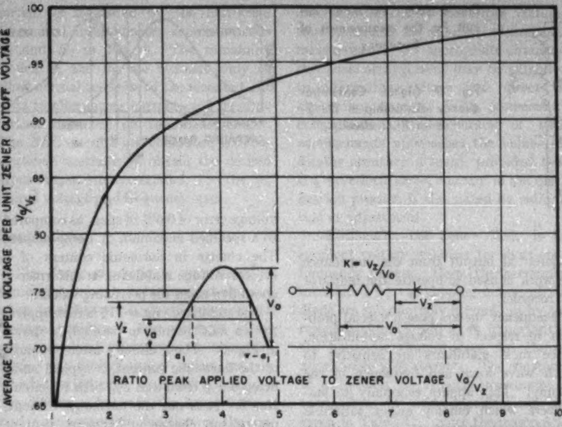


Fig. 5. Relationship between average zener voltage and ratio of peak applied to Zener cutoff voltage (equation)

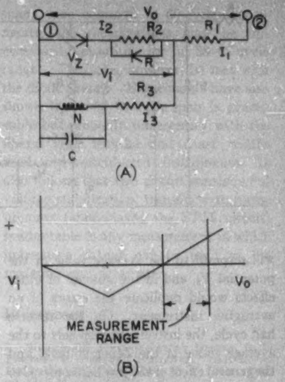


Fig. 6. A—Circuit for limited range voltage measurement. 6. B—Input-output characteristic of circuit in A

urement. The circuit element N represents a zero-left, d-c milliammeter which together with resistor R_3 and capacitor C, measures the average d-c component of the voltage V_i. It is evident that this component will be the difference of the forward and reverse half cycles of V_i as a sinusoidal voltage V_0 is applied to the circuit. Since the forward rectifier voltages will cancel each other, we have with terminal 1 positive:

$$V_i = \frac{V_0}{1.11}\left(\frac{R_2R_3}{R_1(R_2+R_3)}\right) + R_2R_3$$

where V_i = average half cycle voltage, V_0 = half-cycle rms voltage, and other quantities are as shown in Fig. 6(A).

On the reverse half cycle $V_i' = V_a$ and the net d-c component is

$$V_i' = \frac{V_0}{1.11}\left(\frac{R_2R_3}{R_1(R_2+R_3)+R_2R_3}\right) - V_a$$

This relationship, plotted in Fig. 6(B) shows that the instrument will indicate zero at zero voltage, deflect down scale for voltage values below the measurement range, return to zero at the beginning of this range and defect to full scale at maximum voltage.

This circuit was experimentally evaluated for the following application:

Measurement range	110–130 volts
Frequency	60 cycles
Output instrument	9-inch rectangular switchboard 1-milliampere milliammeter
Scale length	14.2 inches
Volt-ampere input	10 maximum

The results were, in general, satisfactory. The only nonlinearity in the circuit, that of the Zener characteristic, causes a small expansion of the scale distribution at higher voltages. This is advantageous rather than the reverse, since it facilitates a scale design which avoids ambiguity of the zero with the initial scale interval. Specifically, normal voltage is marked at the 12-o'clock position, then equal voltage increments on either side will locate the minimum voltage slightly above zero as shown in Fig. 7.

Since the output instrument has an accuracy rating of 1% of span, or 0.1 volt in this instance, the only design problems are optimization of the input-output characteristic and elimination of or compensation for sources of error. Excellent temperature performance can be achieved by use of low-voltage (6–8 volt) Zener diodes and compensation for the residual error by a thermistor in parallel with a small portion of the resistor R_1. The thermistor is chosen to cause a slight increase in the forward current (terminal 1 positive) of such magnitude as to cancel the effect of the positive Zener coefficient. If several, say 4, Zener diodes are used, it is advisable to by-pass them with a duplicate of the reverse current diode, thus eliminating the cumulative effect of the forward Zener resistances. While relatively exact compensation can be achieved in any one instrument, component variation would limit the effectiveness of compensation without individual adjustment. It is probable that a coefficient of ±0.03%/C can be maintained, as compared to the American Standards Association (ASA) *C39.1* value of 0.075%/C.

The frequency effect at 60 cps (cycles per second) is negligible with no discernible error resulting from a ±10% change in frequency.

The wave-form effect is appreciable but notably less than that of average reading instruments calibrated in rms voltage values. On the forward (up scale) half cycle, it is evident that the instrument

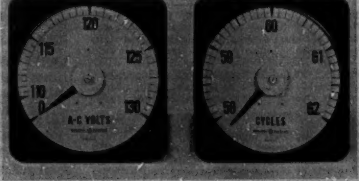


Fig. 7. Switchboard instruments for limited-range measurement of voltage and frequency using Zener diodes

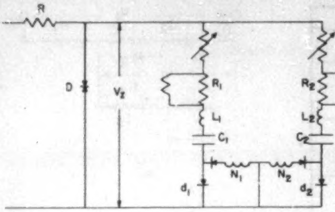
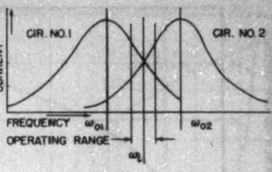


Fig. 8 (left). A diode circuit for the measurement of frequency

Fig. 9 (right). Current-frequency relationship in the 2-series RLC circuits



will respond to the average value of the potential V_0, and in the absence of other effects would duplicate the errors of an averaging instrument. On the reverse half cycle, the instrument responds to the average value of the Zener voltage, and the general effect of the low harmonics is to increase or reduce the Zener average at a greater rate than the forward voltage is changed. Exact cancellation of the wave-form error is, however, obviously impracticable for more than one harmonic, and the Zener voltage was selected to minimize the effect of the third harmonic. Using a ratio of Zener-to-peak voltage of 0.15, the maximum error due to 10% of third harmonic was +0.8% of reading as compared to an error of −3.8% for an average reading instrument. Additionally, filtering will further minimize the wave-form effect but with adverse effects on the frequency characteristic. For this reason, filtering should be limited to applications where frequency is precisely maintained and where the harmonic content is large and variable.

Frequency Measurement

Since frequency meters are inherently limited range instruments, it was not expected that Zener diodes would directly contribute to sensitivity. They are quite useful, however, in their normal function of voltage stabilization, which is essential to meeting ASA accuracy limits when a differential output from the impedance network is used to operate the indicator or recorder.

Table I. Change in Harmonic Content with Voltage

Input Volts, Rms	Output				
		Third Harmonic		Fifth Harmonic	
	Volts	Volts	%	Volts	%
108	24.0	6.9	29	3.1	13
120	24.3	7.2	30	3.5	14.5
132	24.6	7.4	29.6	3.7	14.8

Conditions:
60-cycle sine-wave input
28.8-volt (average) flat-topped output at 16.6 (average) milliamperes
Two sets of three 8-volt diodes

Frequency meters pose a special problem in respect to voltage stabilization, since most stabilizers are sensitive to both wave form and frequency of the supply. This applies especially to stabilizers which employ either saturable transformers or reactors as the voltage sensitive element. The frequency effect takes the form of: Volts out=(Frequency)^n where n is between 1 and 2. The effect can be calibrated out, but also must be designed out since it introduces an undesirable skew in the scale distribution. However, in the author's experimental work with two forms of saturable transformers, it was observed that the harmonic content of the output increased with the voltage even though the average value of the output voltage remained constant. Since the frequency meter necessarily includes two circuits whose relative admittance to these harmonics is not the same, the frequency meter showed a positive variation with increasing voltage, even at midscale frequencies. After brief experiments with filters, it was concluded that such stabilization was inconsistent with our objectives of simplicity and flexibility of application. In particular, it would be necessary to develop stabilizers for each required voltage and frequency range.

Attention was then turned to Zener-diode stabilization using the circuit shown in Fig. 8. Here the incoming voltage wave is clipped to a constant average voltage V_z by the back-to-back group of diodes D. Both the voltage and current of the frequency meter must be held to a minimum to limit the number of diodes used. The output voltage wave is trapezoidal in shape, having about 30% third- and 15% fifth-harmonic content. However, the harmonic content does not change appreciably with voltage as given by the figures in Table I.

A change of ±10% in primary voltage causes a change of $\pm 1^1/_4$% in secondary fundamental voltage, with a calculated voltage error of 0.6% of span, as compared to a specified maximum of 1.5% of span. The change in harmonic content of a typical voltage stabilizing transformer is about five times the preceding values.

The stabilized voltage V_z is then applied to two RLC circuits in parallel, as shown in Fig. 8. These circuits filter out most of the harmonic content, a typical value being 6.9% third and 2% fifth for circuit No. 2. It is not the harmonic content, per se, but change in harmonic content that causes abnormal voltage error with saturable-type stabilizers.

The frequency-current characteristics of the RLC circuits are shown in Fig. 9. Circuit no. 1 is resonant below, and circuit no. 2 is resonant above the operating range with approximate linearity of the differential current in this range. This

Table II. Constants of Type AB-18 120-Volt 58–62-Cycle Frequency Meter

		ASA Specifications	Type AB-18
Rating; volts		120	120
Frequency; cycles per second (cps)		58–62	58–62
Maximum input current at 120 volts 60 cps; milliamperes		50	19
Maximum volt-ampere consumption at 120 volts 60 cps; volt-amperes		6.0	2.3
Scale distribution			linear
Rated accuracy,	% span	3.0	3.0
	cycles	0.12	0.12
Damping factor		1.5 minimum	2.5
Response time, seconds		5 maximum	2.5
Self heating errors, 6 hours at 120 volts, 60 cps			negligible
Temperature influence at 60 cps, ±10 C; % span:			
Uncompensated		1.0	1.75 maximum
Compensated		1.0	0.25 maximum
Effect of 65 C* at 60 cps (% span):			
Uncompensated		4.0	6.0
Compensated		4.0	2.5
Voltage influence ±10% voltage change; % span		1.5 maximum	0.75 maximum
Wave-form influence—effect of 10% third plus 10% fifth harmonic peak on peak; % span		not specified	0.25 maximum

* Effect of 65 C is not included in ASA specifications. The Navy specification tolerance is used instead.

current is measured by the half-wave differential milliammeter shown as coils N_1 and N_2 in Fig. 8. The remaining diodes, d_1 and d_2, are included only to allow normal response of the resonant circuits to alternating current, and have no further influence on the measurement. The *RLC* circuits are computed by conventional methods to obtain the desired input-output characteristics, for the required voltage and frequency span.

Conclusions

The Zener diode is a component of significant value in measurement circuits, when as either a voltage reference or as a voltage stabilizer. One example of its use as a voltage reference is the expanded scale, or limited-range voltmeter, in which the indicator essentially measures the difference between the Zener voltage and an appropriate fraction of the unknown voltage. The basic indicator accuracy is then applicable to this voltage difference instead of the entire voltage, thus enhancing the accuracy of the total measurement. By appropriate circuitry, this basic arrangement may be extended to alternating voltage and direct or alternating current. With temperature compensation, the accuracy of such arrangements approaches the value (indicator accuracy x span), provided that the wave-form error, inherent in the rectification process, is eliminated by calibration or adjustment.

Additionally, the Zener diode is an excellent voltage stabilizer for short-span frequency meters. Here, the harmonics introduced by the clipping action of the diode are substantially filtered out by the resonant circuits, and more importantly, the harmonic content does not change appreciably with current. The constants of a 120-volt 58–62-cycle instrument in comparison to ASA specifications are given in Table II. Since the major ASA tolerances are given in per cent of span and were written when 8- to 10-cycle spans were common in 60-cycle instruments, the specifications may be unnecessarily exacting for a 58–62-cycle range. They can, however, be met with the diode circuit. Experiments have also shown that a 59–61-cycle span is practicable but probably unnecessary with the liberal scale lengths and chart widths used on switchboard instruments. It also follows that this circuit consisting of voltage stabilization, transduction, measurement (abbreviated the STM circuit) is adaptable to any measurement in which impedance change can be used as the measurement parameter.

References

1. SILICON PN JUNCTION ALLOY DIODES, G. L. Pearson, B. Sawyer. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Nov. 1952, pp. 1348–51.

2. THE SUITABILITY OF THE SILICON ALLOY JUNCTION DIODE AS A REFERENCE STANDARD IN REGULATED METALLIC RECTIFIER CIRCUITS, D. H. Smith. *AIEE Transactions (Electrical Engineering)*, vol. 73, pt. I, 1954 (Jan. 1955 section), 645–51.

A Generalization of the Calculus of Finite Differences to Nonuniformly Spaced Variables

GABRIEL KRON
FELLOW AIEE

Synopsis: In the conventional calculus of finite differences the *n* sets of *k* independent variables $x_{(1)}^{\alpha}, x_{(2)}^{\alpha} \ldots x_{(k)}^{\alpha}$ are assumed uniformly spaced along each of the *k* axes of a rectangular frame. Occasionally nonuniform spacings are also assumed, but only along one dimension.

In the proposed calculus the *n* set of *k* independent variables (representing, say, experimental data) are considered as *n* arbitrary points scattered around in a *k*-dimensional Euclidean space. These points are used as anchors for the construction of a new type of reference frame. The frame consists of a network of branches connecting neighboring points; plus a network of planes connecting neighboring branches; plus a network of cubes connecting neighboring planes; and so on, until a network of *k*-dimensional cubes (*k*-spaces) connect the neighboring $k-1$ dimensional cubes. These *k* networks are also interlinked with each other, forming boundaries of one another.

The dependent variables y_α (one or more sets) and their differences of various orders are no longer considered as electric currents and voltages, but rather as electromagnetic waves propagating across the complex of interconnected 0-, 1-, 2-. . . to *k*-dimensional networks, leaving in their wake electric charges and magnetic poles. These generalized waves obey the generalized Stokes' theorems of combinatorial topology. By subjecting the *k* underlying networks, as well as the superimposed electromagnetic waves, to various manipulations so familiar to electrical and electronic engineers, much of the ingredients of a new nonuniform calculus are available for the study of a wide variety of problems in advanced numerical analysis, statistics, and physics.

Of course, the paper submits only a mechanism, not mathematical proofs of the mechanism. The proofs lie in the physical background of the concepts. This paper extends the complete algebraic diagrams developed for a linear graph, to show the tensorial concepts and transformation theory to be used in connection with the entire complex of *k* networks. A proposed path for the electromagnetic waves, to be assumed in simple *k*-dimensional interpolation problems by Newton's formula, is also shown. Eventual solution of extensive problems in nonuniform *k*-dimensional curve-fitting, also smoothing of experimental data, among other things, is anticipated, with the aid of the nonuniform wave-model and the method of tearing.

Physical Problems

PHYSICAL BACKGROUND FOR SOLVING DIFFERENTIAL EQUATIONS

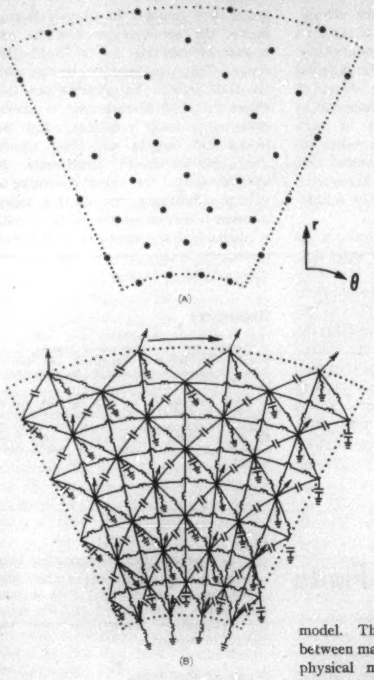
In solving partial differential equations of physical problems (say when analyzing the stresses of an elastic body), it is customary to replace the physical structure by a set of uniformly spaced points in space, at $\Delta x_{(1)}^{\alpha}$, $\Delta x_{(2)}^{\alpha} \ldots$ distance apart, Fig. 1(A). These points define the independent variables $x_{(1)}^{\alpha}$, $x_{(2)}^{\alpha}$, $x_{(3)}^{\alpha}$. Then to each point a dependent variable y_α is assigned, as well as a succession of first-, second-, third-, etc. differences Δy_α, $\Delta^2 y_\alpha$, $\Delta^3 y_\alpha \ldots$. As a result of discretization, a set of algebraic equations (linear or nonlinear) are generated and solved.

The author in his writings always

Paper 58-787, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 7, 1958.

GABRIEL KRON is with the General Electric Company, Schenectady, N. Y.

The author is indebted to his associate, James L. Dolby, for initiating the problem and for numerous conversations and suggestions.



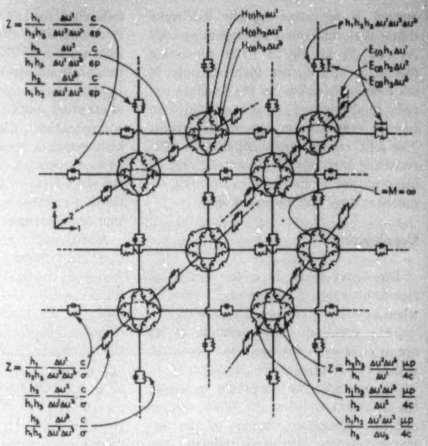

$Z=\frac{h_1}{h_2h_3}\,\frac{\Delta u^1}{\Delta u^2\Delta u^3}\,\frac{c}{\epsilon p}$
$\frac{h_2}{h_1h_3}\,\frac{\Delta u^2}{\Delta u^1\Delta u^3}\,\frac{c}{\epsilon p}$
$\frac{h_3}{h_1h_2}\,\frac{\Delta u^3}{\Delta u^1\Delta u^2}\,\frac{c}{\epsilon p}$
$H_{(1)}h_1\Delta u^1$
$H_{(2)}h_2\Delta u^2$
$H_{(3)}h_3\Delta u^3$
$\rho h_1h_2h_3\Delta u^1\Delta u^2\Delta u^3$
$E_{(1)}h_1\Delta u^1$
$E_{(2)}h_2\Delta u^2$
$E_{(3)}h_3\Delta u^3$
L=M=∞
$Z=\frac{h_1}{h_2h_3}\,\frac{\Delta u^1}{\Delta u^2\Delta u^3}\,\frac{c}{\sigma}$
$\frac{h_2}{h_1h_3}\,\frac{\Delta u^2}{\Delta u^1\Delta u^3}\,\frac{c}{\sigma}$
$\frac{h_3}{h_1h_2}\,\frac{\Delta u^3}{\Delta u^1\Delta u^2}\,\frac{c}{\sigma}$
$Z=\frac{h_2h_3}{h_1}\,\frac{\Delta u^2\Delta u^3}{\Delta u^1}\,\frac{\mu p}{4c}$
$\frac{h_1h_3}{h_2}\,\frac{\Delta u^1\Delta u^3}{\Delta u^2}\,\frac{\mu p}{4c}$
$\frac{h_1h_2}{h_3}\,\frac{\Delta u^1\Delta u^2}{\Delta u_3}\,\frac{\mu p}{4c}$

Fig. 1 (left). Rotating elastic-field structure along cylindrical co-ordinates

A—Network of "difference" equations
B—Network of "pre-differential" equations

Fig. 2 (above). Topological model of the field equations of Maxwell for orthogonal curvilinear co-ordinates

follows a different approach to discretization. He also replaces the physical structure by a set of spatial points. Afterward, however, he interconnects the neighboring points by a set of branches, Fig. 1(B), in which electric currents are assumed to flow. The pattern of branches, and the impedances assigned to them, depend on the nature of the differential equations. The currents and voltages, superimposed upon the branches, represent the dependent variables, their first (and sometimes second) differences, and the divided differences. Mathematicians usually introduce differences Δy_α with order much higher than there are differentials dy_α in the given equations (in order to introduce better accuracy). However, the author never indulges in that practice. He reasons that there are no physical bases behind any differential coefficients which are not included in the given differential equations themselves.

The emphasis is on the word "physical," which also implies "geometrical" or other conceptually-existing model. The author always differentiates between mathematical manipulations and physical manipulations. He aims at introducing into the solution of his problems (if an introduction is possible), only such mathematical manipulations as have a physical (or geometrical) interpretation. Thus he always attempts to set up a physical model for the problem at hand, in the form of an electric circuit, before starting its solution.

ENDOWING EQUATIONS WITH AN UNDERLYING "STRUCTURE"

The physical models used systematically by the author offer certain advantages that purely mathematical equations do not imply. For instance, in partial differential equations the model corresponds to the physical picture that mathematicians have established for a finite material block $\Delta x\Delta y\Delta z$, before going to the limit to establish the partial differential equations. As another example, the electric-circuit models of a mechanical structure give easy access to all the internal stresses and strains that reside in the system.

Moreover, because of the availability of a model, it is comparatively simple to solve large physical systems in small installments by tearing the system apart into an arbitrary number of small structures and solving each separately. The constraint forces appearing at the points of tear are available outright by inspection, without any search.

In general it may be said that the purpose of an electric-circuit model is to endow the formless, amorphous equations (the "algebraic structure"), with an underlying "topological structure." The latter keeps the algebra within proper

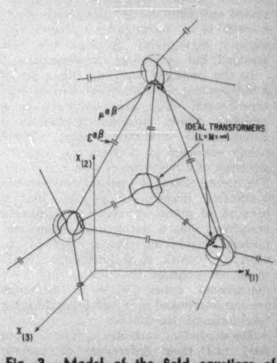

$\mu^{\alpha\beta}$
$\epsilon^{\alpha\beta}$
IDEAL TRANSFORMERS (L=M=∞)
$x_{(2)}$
$x_{(1)}$
$x_{(3)}$

Fig. 3. Model of the field equations of Maxwell for nonuniformly spaced points

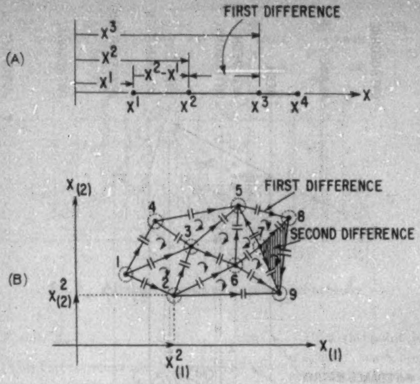


Fig. 4 (left). Independent variables and their differences

A—One independent variable X ($n=4$, $k=1$)

B—Two independent variables $X_{(1)}^{\alpha}$ and $X_{(2)}^{\alpha}$ ($n=9$, $k=2$)

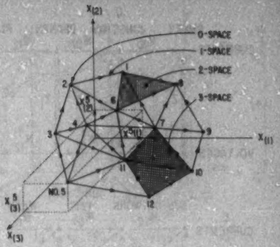


Fig. 5 (above right). Three independent variables ($n=12$, $k=3$)

bounds and directs its course of manipulation and solution.

Of course, the establishment of a structure or model for a physical system is comparatively simple. Often the system itself is its own model (as in electrical networks). In mechanical and dynamical systems, however, it is necessary to endow the underlying topological structure with a superimposed tensor structure in place of a scalar one, before the process of solution begins.

In economic problems the underlying structure is usually hardly discernible, or is entirely absent. It may thus be said that the primary purpose of the theories of economics, of statistics (and of numerical analysis) is to take a given set of amorphous, uncoordinated data and to "breathe life" into them; to endow the numbers with some organization, some "structure." This present paper is an attempt to do just that; namely to set up an underlying topological structure, as well as a superimposed algebraic structure, for a set of arbitrary numbers. The only assumption made about the given set of numbers, is that they belong to an unknown function that possesses derivatives of appropriately high orders at the given points.

ELECTRIC-CIRCUIT MODEL OF THE FIELD EQUATIONS OF MAXWELL

One frequently heard objection against an electric-circuit model of a physical system is that the electric circuit is a linear graph only; an electric circuit cannot represent geometrical planes and cubes with which actual physical systems abound. Since the development to be represented in this paper is based just upon these allegedly missing elements, it becomes necessary to call attention to certain properties of electric circuits so often overlooked.

It is well known that Maxwell, Lord Kelvin, and many other physicists have attempted for many years to establish a mechanical model for the electromagnetic field. All trials ended in failure, because gyroscopes do not behave as displacement currents do. However, by using capacitors, inductors, and ideal transformers, all stationary elements, the author succeeded in establishing a correct (though discrete) electric-circuit model for the electromagnetic field.[4–7] The model shown in Fig. 2 is physically realizable and it has been used to solve steady-state wave-guide and other problems (in two dimensions) by many independent workers. The model is also correct for 3-dimensional transient wave-propagation studies.

In the model of Fig. 2 the copper wires carrying the electric currents do form a linear graph; but the magnetic flux lines linking the wires do not; nor do the electrostatic flux-lines passing between the plates of a condenser. The flux lines pass through planes (bounded by wires), and also pass through cubes (bounded by planes and wires). The cubes may or may not be filled with electric charges and magnetic poles.

The most gratifying, and of course crucial, feature of the model is the appearance, in the ideal transformers, of arbitrary number of magnetic flux lines to represent the arbitrary nature of the magnetic vector-potential *A*. The numbers of these flux lines depend on the entire history of the network model.

The *k*-dimensional models of the present paper are based upon the circuit of Fig. 2. First the latter circuit is generalized from orthogonal curvilinear reference frames to arbitrary reference frames having nonuniformly spaced points, as shown in Fig. 3. The step from three to an arbitrary number of *k*-dimensions requires only geometrical and mathematical extensions and manipulations.

It is well known that Maxwell's field equations satisfy Stokes' theorem (multiple integral relations between divergences and curls of tensors and tensor-densities of various valence). A topological complex of networks, as well as the algebraic structure superimposed upon the underlying structure, also must satisfy Stokes' theorem generalized to *k* dimensions.[8,9] Thus the author feels confident that by allowing his networks and the electromagnetic waves superimposed upon his networks to follow the generalizations for the field equations of Maxwell, both the networks and the waves automatically satisfy also Stokes' theorem.

Of course, in practical computational work the model networks need not be established; not even on paper. A correct code can replace all paper models in any number of dimensions. As a matter of fact, advanced texts on combinatorial topology never show even a linear graph, much less a higher-dimensional model.

Numerical Analysis

INDEPENDENT VARIABLES

Translating the given physical background into the language of numerical analysis, let a set of tabular arguments x^{α} (one set of independent variables) be given, Fig. 4(A). If the arguments are plotted along the x axis, the first differences x_2-x_1 or x_3-x_2 may be visualized as the geometrical concept of distance. But the customary second differences x_3-x_1 or x_4-x_2 cannot be visualized geometrically. By logic, the

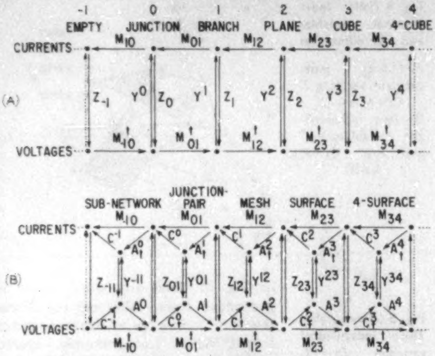

-1
0
1
2
3
4
EMPTY
JUNCTION
BRANCH
PLANE
CUBE
4-CUBE
CURRENTS
VOLTAGES
M10
M01
M12
M23
M34
(A)
Z-1
Y0
Z0
Y1
Z1
Y2
Z2
Y3
Z3
Y4
SUB-NETWORK
JUNCTION-PAIR
MESH
SURFACE
4-SURFACE
(B)
C-1
C0
C1
C2
C3
A0
A1
A2
A3
A4
Z-11
Y-11
Z01
Y01
Z12
Y12
Z23
Y23
Z34
Y34

Fig. 6. Algebraic diagrams

A—Incidence matrices M B—Boundary operators C, A

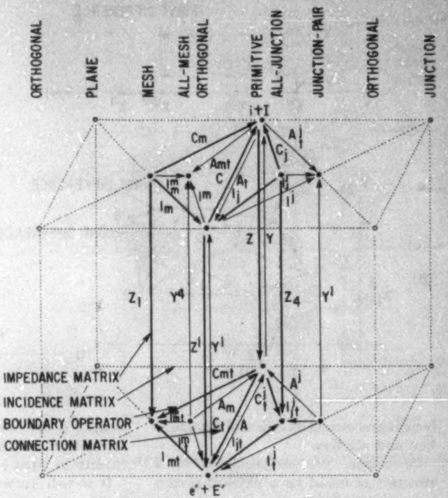

ORTHOGONAL
PLANE
MESH
ALL-MESH ORTHOGONAL
PRIMITIVE
ALL-JUNCTION
JUNCTION-PAIR
ORTHOGONAL
JUNCTION
i+I
Cm
Amt
C
At
Cj
Aj
Im
im
ij
Z
Y
Z1
Y4
Z4
Y1
Z1
Y1
Cmt
Am
Ct
A
Cj
Aj
Imt
imt
ijt
e'+E'
IMPEDANCE MATRIX
INCIDENCE MATRIX
BOUNDARY OPERATOR
CONNECTION MATRIX

Fig. 7 (right). Algebraic diagram of a network with simultaneous mesh and junction excitations

second differences of the independent variables ought to represent areas, if the first differences represent distances. But in Fig. 4(A) they do not. At best, the second differences approximate an area by means of the greatest distance, defined by three neighboring points that lie in one dimension.

With two sets of independent variables $x_{(1)}^{\alpha}$ and $x_{(2)}^{\alpha}$, Fig. 4(B), two points can define a distance and three points an area. But no volumes can be defined. With three sets of independent variables, Fig. 5, lines, planes (triangles), and cubes (tetrahedrons) can be introduced with the aid of two, three, or four neighboring points.

Thus when k independent variables are given, each with n elements, they will be considered as n points in a k-dimensional Euclidean space. The distance between any two points (say 3 and 5) is

$$d_{35}=[(x_{(3)}^{1}-x_{(5)}^{1})^2+(x_{(3)}^{2}-x_{(5)}^{2})^2+\ldots(x_{(3)}^{k}-x_{(5)}^{k})^2]^{1/2} \quad (1)$$

Table I. Given Dependent and Independent Variables

Dependent	Independent			
y_α	$x_{(1)}^{\alpha}$	$x_{(2)}^{\alpha}$	....	$x_{(k)}^{\alpha}$
y_1	$x_{(1)}^{1}$	$x_{(2)}^{1}$		$x_{(k)}^{1}$
y_2	$x_{(1)}^{2}$	$x_{(2)}^{2}$		$x_{(k)}^{2}$
y_3	$x_{(1)}^{3}$	$x_{(2)}^{3}$		$x_{(k)}^{3}$
....	....	....	....	....
y_n	$x_{(1)}^{n}$	$x_{(2)}^{n}$		$x_{(k)}^{n}$

To establish a linear reference frame, every point is connected with about $2k$ neighboring points, whose distances are shortest. Points at "corners" are connected to only about k neighboring points; those along edges to about $k+1$ points, and so on. The aim is to erect a linear reference frame so that the branches emanating from each point should span the entire k dimensional space.

Similarly, a planar reference frame is rigged up with sets of three points (triangles) covering minimal areas. Next the k-dimensional space is spanned with tetrahedrons having minimal volumes, and so on. The manner of assuming the reference spaces depends on the accuracy of calculations wanted. For instance, a plane may be defined by more than three points, etc. The assumption of a variety of reference frames is analogous to the various types of differences (forward, central, etc.) that are assumed in conventional studies.

The square of the distances, areas, and volumes will be assigned as impedances (inverse admittances, or rather, permeabilities μ or ϵ) associated with each network element. In the interconnected structure (called "polyhedron" or "polytope") each even-dimensional network will be constructed of magnetic material and each odd-dimensional network will be constructed of dielectric material.

With k independent variables, it is possible to build up higher than kth differences. But in that case the points can not define larger than k-dimensional volumes. With $k=1$ all results become, then, identical to those of the conventional calculus of finite differences.

To establish divided differences and to interpolate, it will be sufficient to assume that the polyhedron is torn dimensionwise into k isolated networks. Each q-dimensional network itself is, however, left interconnected.

Dependent Variables

Let now a set of dependent variables y_α also exist, in addition to the n sets of k independent variables $x_{(k)}^{\alpha}$, shown in Table I. (These variables can be assumed to be complex numbers, but for the sake of better visualization, they are real.) The variables originate, say, in connection with a set of observations, or experiments.

The dependent variables y_α will be assumed to be represented by magnetic poles. The poles produce B_O magnetic flux lines, placed on their respective points. Because the flux-lines will be assumed to undergo sinusoidal variations in time (say 60 cycles), the y_α also represent generated voltages

$$y_\alpha=B_O=dB_O/dt \quad \textbf{(2A)}$$

that appear upon each junction-point of the complex. Thus at the end-points (boundaries or junction-pairs) of each branch there appears a difference of potential

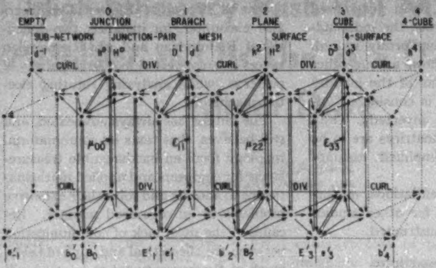


Fig. 8. Algebraic diagram of a sequence of networks

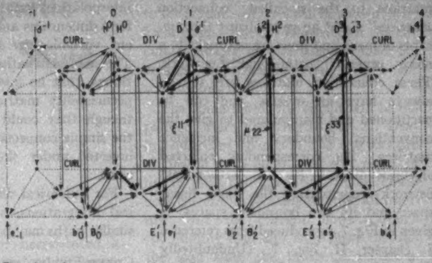


Fig. 9. A sequence of transformation steps to generalize Newton's interpolation formula

$$E_1 = \text{div}\ \dot{B}_O \qquad (2B)$$

(This first step may also be considered by assuming an absolute potential V at each point, $V = y_\alpha$. Then across each junction-pair there appears the difference of potential

$$E_1 = \text{grad}\ V\,) \qquad (2C)$$

Since each linear branch has a dielectric permeability ϵ^{11} proportional to its inverse (length)^2, an electrostatic flux-line D^1 flows in each branch

$$D^1 = \epsilon^{11} E_1 \qquad (2D)$$

Thus the total stored electrostatic energy in the entire linear network is equal to the sum-of-squares of all first-order divided differences

$$U_{(1)} = \Sigma E_1 \epsilon^{11} E_1 = \Sigma D^1 (\epsilon^{11})^{-1} D^1 = \left(\frac{y_n - y_{n-1}}{x_n - x_{n-1}}\right)^2 \qquad (2E)$$

(The "sum-of-squares" concept in statistics is as basic as the "total-stored-energy" concept is in physics.)

By Maxwell's electromagnetic field equations it is known that a sinusoidal $dD/dt = \dot{D}$ is equivalent to a displacement current, and that such a current also produces a magnetomotive force H by the relation

$$\dot{D}^1 = \text{curl}\ H^2 \qquad (2F)$$

However, this H^2 appears upon the planes and not upon the linear branches. Since the planes are assumed to be constructed of magnetic materials having a permeability μ^{22}, which is proportional to the inverse (area)^2, a magnetic flux B_2 appears across the planes

$$B_2 = \mu_{22} H^2 \qquad (2G)$$

where μ_{22} is the inverse of μ^{22}. The total stored magnetic energy in all second-order divided differences.

$$U_{(2)} = \Sigma H^2 \mu_{22} H^2 = \Sigma B_2 (\mu_{22})^{-1} B_2 \qquad (2H)$$

Up to this point one cycle of the electromagnetic wave-propagation has been generated. The production and propagation of E_3, D^3, then H^4, B_4, etc. continues, until the kth dimension is reached. In each network surplus electric charges or magnetic poles accumulate during propagation which may or may not influence the further course of propagation, depending upon the problem at hand. Also the total stored energy in each network is equal at all times to the sum-of-squares of the respective divided differences.

Ideological Graphs

ALGEBRAIC DIAGRAMS

The "algebraic diagram" of Fig. 6 (reproduced from reference 3) shows all transformation tensors C, A and all impedance tensors Z, Y associated with the study of a system, as well as it shows all relations between these tensors. In particular, Fig. 6(A):

1. The dots represent the spaces of currents (upper row) and voltages (lower row). Each dot itself represents an entire network (linear, planar, cubic, etc. network).
2. The horizontal arrows represent the "incidence" matrices M. They show how the basic spaces are related to each other.
3. The vertical arrows represent the impedance and admittance matrices Z and Y associated with each network.
4. Fig. 6(B) shows by slant arrows the manner of forming the boundary spaces associated with the basic spaces. These inter-relations are expressed by matrices C and A, called "boundary operators."

The vertical Z and Y matrices are square and usually non-singular. (They usually have an inverse.) The transformation matrices contain +1, −1 and 0 as their elements, and are characterized the fact that they are rectangular, not square, and thus they have no inverse.

The product of two matrices is represented on the algebraic diagram by the sum of two arrows. The sum is either a third arrow or zero. Thus from Fig. 6(B) (going against the arrows by convention)

$$C^1 A_t^2 = M_{12}$$

$$A_t^3 C^2 = 0 \qquad (3)$$

Moreover the currents are represented on the algebraic diagram by short vertical arrows (not shown on Fig. 6, only on Figs. 8 and 9) entering each space of the upper set of dots. The voltages are similarly represented by short vertical arrows entering the lower set of dots. Thus the algebraic diagram supplies visually all "dead" and "live" tensors associated with the system, and also all possible algebraic manipulations with the tensors. The analysis of a problem consists thereby of picking out, from all possible routes of procedure, the most suitable ones that require the minimum of computations.

ENTER ENGINEERING

As has often been the case in the past, theoretical mathematicians are not interested primarily in solving practical problems. They erect an artistic and self-contained theoretial structure and leave it to the engineers and physicists to make it workable in practice. Similarly the incidence matrices and boundary operators of combinatorial topology, because of their rectangular form, do not allow a free unrestricted passage back and forth between the various spaces. However, in all practical applications it is an absolute necessity to travel to and fro freely in all directions: from currents to voltages; from the equations of state to the equations of solution; from one dimension to another dimension; and so on.

In order to acquire such a freedom of movements, the author has generalized the concepts of incidence and boundary matrices to the so-called "connection matrices" C_β^{α}[2,3] by generalizing the concepts of mesh and junction networks to a more general "orthogonal" network. These connection matrices are square, always have an inverse, and can be partitioned and manipulated to give the conventional boundary and incidence matrices. The corresponding enlarged algebraic diagram for one "orthogonal" network (representing the single linear space and its two boundary spaces) is given in Fig. 7, reproduced from reference 3. chapter II, Fig. 7. (Undoubtedly these nonsingular transformation matrices may have already been recognized by topologists, and are hidden somewhere among their esoteric symbolism and terminology.)

Now, a sequence of such complete algebraic diagrams, given in Fig. 8, shows all the available tensorial concepts and all the possible routes of procedure, for the numerical solution of problems under consideration. The diagram can be further enlarged and generalized, as it has often been done in reference 3 for the linear portion. (There are also other cross-paths not shown on the diagram.)

Practical Applications

Interpolation

The algebraic diagram also enables the engineer to plan his passage through the labyrinth of spaces with the least amount of numerical computations for any particular problem or assumptions. Thus the algebraic diagram serves as a "thinking-graph."

For example, the heavily-drawn arrows on Fig. 9 show one of the several possible steps that can be taken in using Newton's interpolation formula, when generalized for k dimensions and when the independent variables are nonuniformly spaced. The assumed paths show, among other things, that no inversion of the induced permeability matrices is considered (although they could be inverted). Only the simple connection matrices are to be inverted under the simplified assumptions.

For numerical differentiations, numerical integrations, and for other studies similar paths may be constructed.

Curve-Fitting and Smoothing

It should be noted that during the simple wave-propagation assumed for interpolation, numerical differentiation, and integration, etc., purposes, no feedback was assumed to exist among the waves. As soon as more involved statistical or physical problems are investigated (such as curve-fitting, or smoothing), the waves in the various spaces begin to react upon each other and thereby the induced permeability matrices must be inverted, either singly or simultaneously.

The author and his associates anticipate the appearance of linear and nonlinear equations, boundary value and eigenvalue problems, as well as optimizations of various types, in which all spaces influence the answer to various degrees. In the solution of such extensive problems the method of tearing[3] promises to become a useful tool.

Curved Spaces

The Euclidean planes and cubes of higher dimensions are, of course, only the simplest types of building blocks that form a complex. Combinatorial topology assumes almost any imaginable curved spaces as elements. Electrical engineers also encounter, for instance, two-dimensional Riemannian and non-Riemannian spaces interconnected in a large variety of patterns, while analyzing rotating electric machinery.

The author has always advocated, and still believes, that texts on combinatorial topology form an inexhaustible treasure-house for engineers and applied mathematicians to analyze and solve their extensive problems in an organized manner. Because of the utter lack of communication between the theoretical and applied fields, much of the concepts yet to be used will have to be rediscovered, processed, and put to practical use by the interested engineers themselves.

References

1. Foundations of Algebraic Topology (book), S. Ellenberg, N. Steenrod. Princeton University Press, Princeton, N. J., 1952.

2. Tensor Analysis of Networks (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1939.

3. Diakoptics—The Piecewise Solution of Large-Scale Systems, G. Kron. *The Electrical Journal*, London, England, serial beginning June 7, 1957.

4. Equivalent Circuits of the Field Equations of Maxwell—I. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 32, nos. 5 and 6, May 1944, pp. 289–99.

5. Network Analyzer Studies of Electromagnetic Cavity Resonators, J. R. Whinnery, C. Concordia, W. Ridgeway, G. Kron. *Ibid.*, no. 5, pp. 360–67.

6. A New Approach to the Solution of High-Frequency Field Problems, J. R. Whinnery, Simon Ramo. *Ibid.*, pp. 284–88.

7. Models of Electromagnetic Field Analysis, S. A. Shelkunoff. *Bell System Technical Journal*, New York, N. Y., vol. 27, July 1948, pp. 487–509.

8. On Multiple Integrals, G. de Rham. *Hamburger Abhandlungen*, Hamburg, Germany, vol. 12, 1937–38, pp. 313–39.

9. Boundary Theorems of Tensor Analysis in the Large, A. W. Tucker. *Pamphlet*, Pasadena, Calif., June 1941.

ERRATUM

"A Stabilized D-C Differential Transistor Amplifier" by Louis Depian and Richard E. Smith, published in Communication and Electronics, May 1958, pages 157-59.

On page 159, the captions, "Fig. 7. Differential amplifier with self-bias" and "Fig. 8. Modification of differential amplifier with self-bias" should be interchanged. Also, the unlabeled resistor in Fig. 8 should be labeled: R_6.

High-Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive

C. L. BOYAJIAN
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a d-c push-pull magnetic amplifier which is used to drive a d-c shunt motor. A circuit called the high-efficiency push-pull magnetic amplifier circuit was used successfully to drive a 1/20-horsepower d-c shunt motor.[1]

This circuit has made the application of magnetic amplifiers to driving d-c motors in reversible operation practical.

The general operating characteristics of this circuit are discussed and actual data presented. The problems associated with driving a d-c motor load and the factors influencing the design of the circuit are discussed. The results achieved in using the high-efficiency circuit in an actual application are presented.

A COMMONLY used d-c push-pull full-wave magnetic amplifier circuit is shown in Fig. 1. This circuit has a maximum theoretical efficiency of 17%. This circuit has a maximum theoretical efficiency of 17%. This efficiency is based on the assumption that the internal impedance of the amplifier circuit as seen from the load is negligible and that the load resistance is equal to the $\sqrt{2}$ times the resistance of one of the dummy load resistors.

The large power loss associated with this circuit is undesirable of course. This becomes of greater concern as the power level of the application increases.

A different magnetic amplifier circuit was invented in 1952 by B. D. Bedford of the General Electric Company. This amplifier circuit has a maximum theoretical efficiency of 50% as compared to the 17% efficiency of the circuit of Fig. 1. The circuit was called the high-efficiency push-pull circuit because of its relatively higher efficiency.

Before the invention of the high-efficiency circuit, it was not too practical to drive a d-c motor in push-pull operation with a magnetic amplifier. The excessive loss of power necessitated a magnetic amplifier considerably larger than the motor load. The high-efficiency circuit makes possible an appreciable reduction in the size of a magnetic amplifier for a given load.

Basic Operating Characteristics

The high-efficiency push-pull full-wave d-c magnetic amplifier circuit is shown in Fig. 2. To understand the basic operating characteristics of the high-efficiency circuit, resistive loads will be considered first.

This circuit is the equivalent of four simple magnetic amplifier bridge circuits connected together as a bridge. The actual load is connected in the center of the bridge arrangement. The dummy load R_z is connected externally as shown. Only four cores are necessary, as shown in Fig. 2, since two gate windings are placed on one core.

The dummy load performs the following functions in the circuit.[2] First, the circuit would not work properly without it because the dummy load provides a reverse voltage across the resetting cores. The reverse voltage which prevents current flow through the diodes is necessary for the core flux to reset. Second, the dummy load impedance is necessary to help limit the standby current of the amplifier.

Maximum theoretical efficiency will result when the dummy load is made equal to the actual load. As will be shown later, for a given load, the efficiency of the circuit decreases as the dummy load decreases, approaching zero output for zero dummy load.

The operation of the circuit will be explained briefly referring to Fig. 2. During quiescent operation, the reactors are all biased to a point that will result in linear push-pull operation. The quiescent current is the same through all windings. All the current flows through the dummy load and none flows through the load since the bridge is balanced. When a d-c signal of a given polarity passes through the control windings, the cores of one reactor are turned in the "on" direction while the cores of the other reactor are turned in the "off" direction. This causes the bridge to be unbalanced, resulting in a full-wave rectified voltage of a corresponding polarity to appear across the load. A reverse signal will reverse the dummy load is always of the same polarity.

The resistors shunting the diodes provide a degenerative feedback which stabilizes the gain of the amplifier.

To show the operating characteristics of an actual high-efficiency circuit, three cases were chosen using resistance loads. The dummy load resistance was made large compared to the resistance of the other amplifier components such as the core windings, transformer windings, and diodes. The dummy load was 500 ohms for all three cases. For each case, a plot of load voltage and current, and dummy load voltage and current versus control current is presented. Only the characteristics for one polarity of signal current are shown. For the reverse polarity the load voltage and current reverse but the dummy load voltage and current are of the same polarity.

In the first case, the load resistance was made equal to the dummy load resistance. The resulting characteristics are shown in

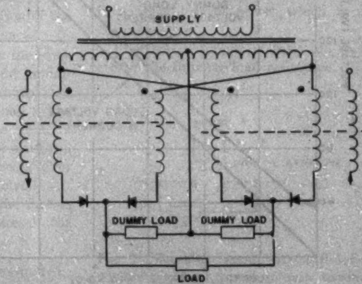


Fig. 1. Push-pull magnetic amplifier

Paper 58-775, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 1, 1958.

C. L. BOYAJIAN is with the General Electric Company, Schenectady, N. Y.

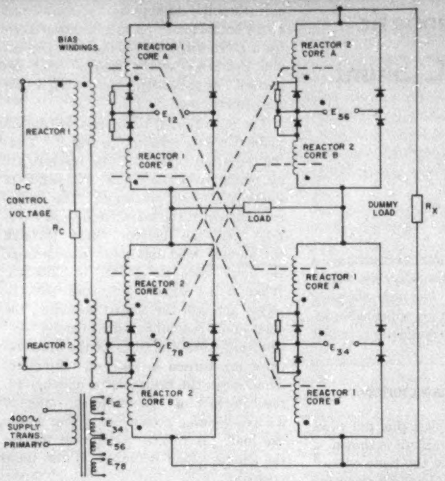

BIAS WINDINGS
REACTOR 1 CORE A
REACTOR 1
E$_{12}$
REACTOR 1 CORE B
D-C CONTROL VOLTAGE
R$_C$
REACTOR 2 CORE A
E$_{56}$
REACTOR 2 CORE B
DUMMY LOAD
R$_X$
LOAD
REACTOR 2 CORE A
REACTOR 1 CORE A
REACTOR 2
E$_{78}$
E$_{34}$
REACTOR 2 CORE B
REACTOR 1 CORE B
400~ SUPPLY TRANS. PRIMARY
E$_{12}$
E$_{34}$
E$_{56}$
E$_{78}$

Fig. 2. High efficiency push-pull full-wave d-c magnetic amplifier circuit

Fig. 3. From this plot, it can be seen that the maximum theoretical efficiency of 50% is approached.

In the second case, the load resistance was made five times the dummy load resistance. Fig. 4 shows that the voltages are about the same as they were in Fig. 3. The voltage across the load did not become greater than the voltage across the dummy load as might be expected.

Actually, the load voltage cannot exceed the dummy load voltage. The instant the load voltage exceeded the dummy load voltage, the complementary diodes associated with the "off" reactor would conduct. This bypasses the current from the load to the dummy load until the two voltages are equal. This is why the dummy load current is greater than the load current even in saturation, when one reactor is turned completely off. Again this shows why the dummy load is necessary.

In the third case, the load resistance is made 1/5 the dummy load resistance. Fig. 5 shows that for this condition the voltages divide approximately as the ratio of the resistance as would be expected. The current is therefore approximately the same. Initially, the load voltage has a greater slope, which is approximately the same as when the voltages divided equally as in the first and second cases. This results in the load current increasing at a rate that would make it exceed the dummy load current. However, this is not possible since any current that passes through the load must also pass through the dummy load. The dummy load restricts the level of load current resulting in the two currents being approximately the same. In some applications, a nonlinear gain characteristics such as this is undesirable.

The data shown in Figs. 4 and 5 show that the efficiency of this circuit decreases whether the load is greater or less than the dummy load.

The data in Fig. 6 show the push-pull operation of the circuit. These data will be used to compare the results achieved when the motor load is used.

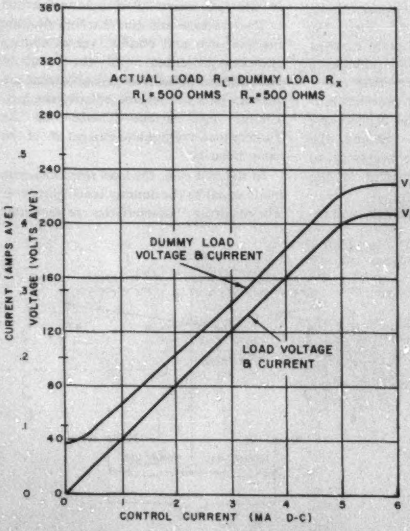
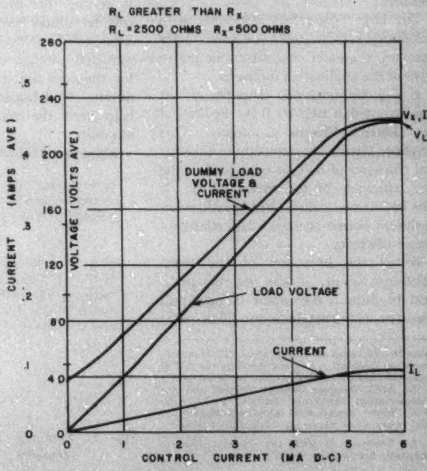
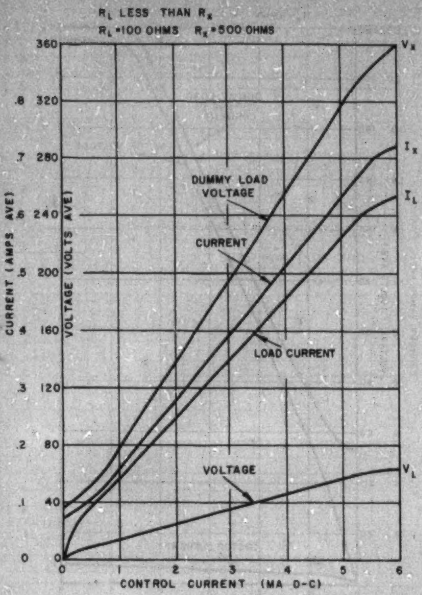


Fig. 3 (left) and Fig. 4 (below). Amplifier voltage and current characteristics for resistance loads





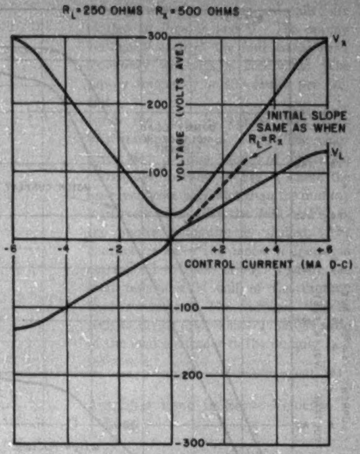


Fig. 5 (left). Amplifier voltage and current characteristics for resistance loads

Fig. 6 (above). Amplifier push-pull voltage characteristics

A summary of the characteristics of this circuit as shown in Figs. 3, 4, and 5 is as follows:

1. V_L can not exceed V_x.
2. I_L can not exceed I_x.
3. When $R_L > R_x$, the approximate characteristics are, $V_L = V_x$ and $\frac{I_L}{I_X} = \frac{R_X}{R_L}$.
4. When $R_L < R_X$, the approximate characteristics are $I_L = I_X$ and $\frac{V_L}{V_X} = \frac{R_L}{R_X}$.
5. Maximum efficiency is achieved when $R_L = R_X$.

Motor Load

When the high-efficiency circuit is used to drive a d-c shunt motor, there are a number of factors to be considered. For this discussion, the motor field is separately excited so that the amplifier load is the armature of the motor.

The effective impedance of the motor varies considerably for different modes of operation. These modes are, the steady state conditions of stalled motor and motor running at a constant velocity, and the transient conditions of acceleration and deceleration. This presents the problem of matching the dummy load to the motor load for highest efficiency or deciding what dummy load will result in the best over-all operation for a particular application.

When the motor is stalled, the armature presents a resistive load to the d-c output of the amplifier. When the motor is running at a constant velocity with no load, the armature presents a high effective-load impedance, relative to the stalled resistance, due to the counter electromotive force voltage generated by the motor armature. This effective impedance was approximately 50 times greater than the stalled armature resistance for the motor employed.

When load is applied to the running motor, the effective impedance decreases. Also, transient operation of the motor presents varying impedances.

Because of these variations in the effective motor impedance, it would be impractical to try to design a dummy load that would vary as the motor load. The simplest solution is to choose a value of resistance for the dummy load that will result in the best over-all operation for the particular application involved.

The following are some of the factors that would influence the choice of the dummy load resistance:

1. The maximum stalled torque required.
2. The acceleration required.
3. The power dissipated in the dummy load resistor.
4. The duty cycle of the motor.

If the dummy load resistance is made equal to the stalled-motor armature-terminal resistance, maximum stalled current would flow resulting in maximum stalled torque and maximum acceleration for a given supply voltage. The disadvantage of this condition is that when the motor is running with little or no load, the efficiency of the amplifier drops off considerably. This is because the armature current is very low while the dummy load current remains about the same as for the stalled condition since the voltage across it remains the same. This results in a large loss of power which is undesirable.

For this reason, it is desirable to increase the dummy load resistance. The

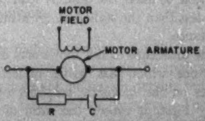


Fig. 7. D-c shunt motor with resistance-capacitance network across armature terminals. Motor field separately excited

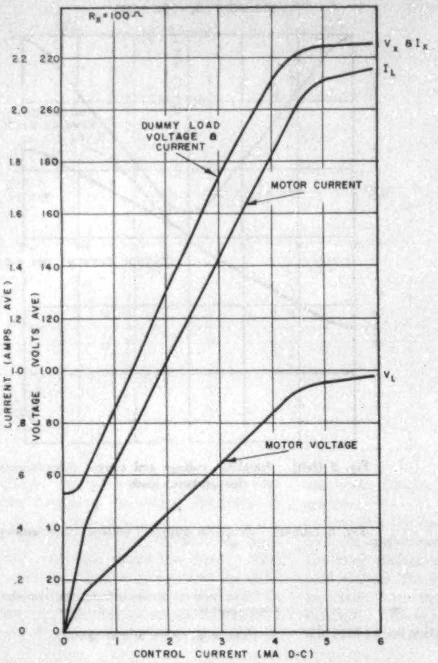


Fig. 8. **Amplifier characteristics with stalled motor load**

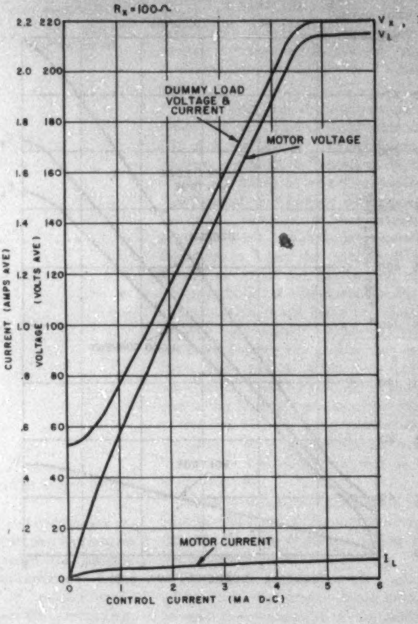


Fig. 9. **Amplifier characteristics with running motor load (no load on motor)**

amount it should be increased depends on how much of the time the motor would be running, especially at light loads, and also on the maximum torque requirements. As the dummy load resistance is increased, the maximum available torque is decreased since the load current is limited by the dummy load resistance as shown in Fig. 5. Increasing the dummy load resistance will decrease the quiescent power dissipation which is desirable. The nonlinear load characteristic shown in Fig. 5 may or may not be tolerated, depending on the application.

Under no circumstances would it be desirable to make the dummy load resistance less than the stalled motor armature resistance.

The armature of the motor has an inductance which has some undesirable effects on the magnetic amplifier operation. First, it causes the ratio of load voltage to dummy load voltage, under stalled conditions, to be lower than it would be if the inductance was zero. Also, the magnetic amplifier investigated had a slight tendency to snap and had a discontinuous gain characteristic.[1,2]

These inductive effects were removed by the addition of a series R-C network across the motor armature terminals as shown in Fig. 7. The network improved the ratio of load voltage to dummy load voltage under stalled conditions to approximately the expected value as shown in Fig. 8. The network also filtered the ripple current through the motor armature resulting in the quiet operation of the motor.

The other factor to be considered is the effect that the magnetic amplifier has on the response of the motor. The motor response depends on the effective impedance of the amplifier as seen from the motor armature terminals. In calculating the motor inertia time constant $T = JR/K_tK_e$, the R becomes the sum of the motor terminal resistance and the effective output resistance of the magnetic amplifier. In the actual application to be discussed next, the motor time constant was approximately doubled.

Actual Data of Amplifier Driving a Motor Load

A presentation of data taken with the magnetic amplifier driving a d-c shunt motor will show the results that can be achieved. The circuit of Fig. 2 was used with the motor load plus the R-C network as shown in Fig. 7. The amplifier is the same as the one used to take these data for Figs. 3–6. The d-c field of the motor was separately excited.

The motor has a continuous rating of 1/20 horsepower at 4,000 rpm and 115 volts. It can deliver a maximum torque output of 50 inch-ounces with an armature current of 2 amperes. The armature terminal resistance is approximately 45 ohms.

Although the continuous rating of the motor is 1/20th horsepower, the amplifier is capable of driving the motor to deliver a peak power output of 1/7th horsepower.

The application for which this circuit was developed has the following requirements. The amplifier must be able to maintain torque saturation of the motor

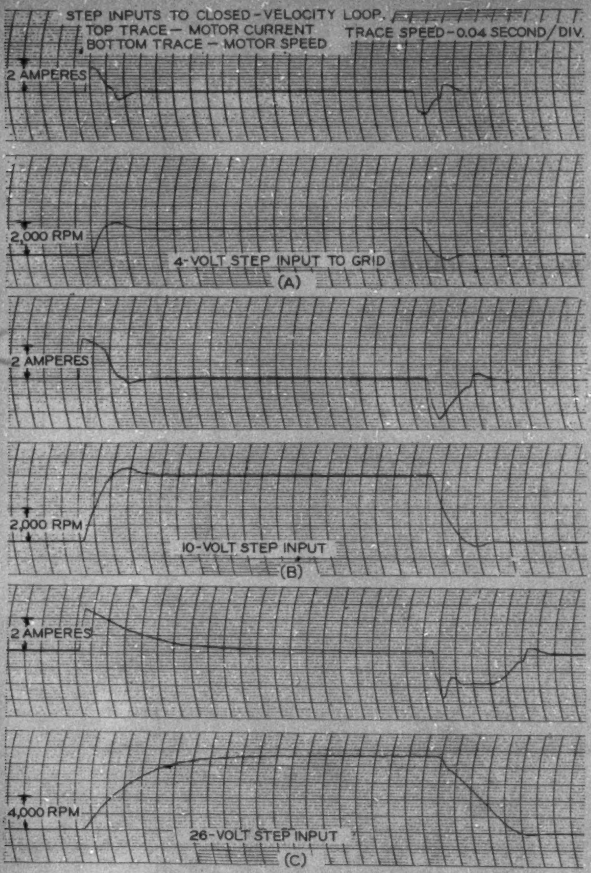


Fig. 10. Transient response of closed velocity servo loop showing motor armature current and motor speed simultaneously against time

until it reaches rated speed to achieve maximum acceleration with step inputs. In other words, the amplifier must deliver 2 amperes to the motor armature until rated motor speed is reached when a large input signal is applied suddenly to the amplifier. The motor during another mode of operation, accelerates and decelerates 50% of the time and runs at a constant velocity with a light load 50% of the time. Also, it is necessary to keep the power dissipation at a minimum because it is to be used in an air-borne application.

Although the motor is to run at a constant velocity with a light load 50% of the time, it was necessary to keep the resistance of the dummy load low to be able to achieve the required high acceleration requirements for step inputs. A dummy load resistance of 100 ohms was used as a compromise between the desired 45 ohms for high acceleration current and the high resistance desired for constant velocity operation, which is about 2,500 ohms.

The data showing the steady state characteristics of the magnetic amplifier driving the motor running and stalled are plotted in Figs. 8 and 9. The supply voltage of each of the four transformer secondary windings is 250 volts. The supply frequency is 400 cycles per second.

For the data in Fig. 9, the motor was running with no load. The characteristics of this mode of operation correspond to those of Fig. 4 where the load resistance is greater than the dummy load resistance. For the data of Fig. 8, the motor armature is stalled. The characteristics of this mode of operation correspond to those of Fig. 6 where the load resistance is half of the dummy load resistance. The voltages can be seen to divide approximately as the ratio of the load resistance to the dummy load resistance.

Amplifier Used in Servo-Velocity Loop

The particular amplifier and motor mentioned were used as part of a servo-velocity loop. The magnetic amplifier was driven by a d-c push-pull electronic driver stage. The resulting time constant of the magnetic amplifier was approximately 0.005 second. The inertia time constant of the motor when driven by the magnetic amplifier was 0.08 second and the armature inductive time constant was approximately 0.001 second. With the proper amount of velocity feedback, the closed velocity loop frequency response was attenuated 3 decibels at an angular frequency of approximately 50 radians per second.

The recorder traces of Fig. 10 illustrate the transient response of the closed-velocity servo-loop system. The step input of Fig. 10(A) was small so as not to saturate the amplifiers. This shows the true response of the servo. The step input of Fig. 10(B) caused the magnetic amplifier to saturate initially resulting in torque saturation of the motor. The steady-state speed of 4,000 rpm is the rated speed of the motor. The step input of Fig. 10(C) resulted in the continuous saturation of the magnetic amplifier.

Conclusions

The high-efficiency push-pull magnetic amplifier circuit has made the application of magnetic amplifiers to driving reversible d-c shunt motors practical. The use of this circuit makes possible a considerable reduction in the size of the magnetic amplifier for a given application compared to the size that would be necessary if the more common push-pull circuit of Fig. 1 were used. This applies mostly to applications where the power level is such that efficiencies become quite important and the increased circuit complexity is justified by the decrease in power lost and the decrease in the size of the amplifier.

References

1. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 431–54, 320–40.

2. Magnetic Amplifiers for the Control of Servomotors, Allen R. Perrins. "Conference on Magnetic Amplifiers," *AIEE Special Publication T-86*, July 1956, pp. 246–54.

3. Inductive Load Instability in Magnetic Amplifiers, H. I. Leon, Allen B. Rosenstein. *AIEE Transactions*, vol. 74, pt. I, March 1955 pp. 8–20.

4. Speed Control of a D-C Motor Using a Magnetic Amplifier, W. Leonhard. *Ibid.*, vol. 76, pt. I, May 1957, pp. 112–19.

5. Magnetic-Amplifier Control of D-C Motors, A. Kusko, J. G. Nelson. *Ibid.*, vol. 74, pt. I, July 1955, pp. 326–33.

6. Push-Pull Magnetic Amplifier, W. R. Seegmiller. *U. S. Patent 2,770,769*, May 1952.

Elimination of Magnetic-Amplifier Control Circuit Inductance

D. P. CHANDLER
ASSOCIATE MEMBER AIEE

R. W. DOWNING
ASSOCIATE MEMBER AIEE

RESPONSE of magnetic amplifiers to signal changes is typically impeded by the necessity to change the average flux linkage of the control winding in order to obtain a different output. This inductive effect requires an elapse of time during which a voltage-time integral, proportional to the required change in average flux linkage, must be applied to the control winding before a new steady-state operating point can be reached.

In this paper the concept of magnetic amplifiers with noninductive control windings is introduced. This concept is based on proper manipulation of well-known phenomena so as to eliminate the requirement for the control-winding flux linkage to change when control and load currents are changed. The possibility of obtaining instantaneous response is demonstrated, and a practical method of approaching the condition of a noninductive control winding is presented for cases of fixed load impedances.

Nomenclature

B_0 = average (d-c) flux density
B_m = amplitude of alternating flux density
H_0 = average magnetic intensity
E = supply voltage, rms
$\mathbf{E}$ = supply voltage vector
E_{AC} = alternating voltage across magnetic-amplifier load winding, rms
$\mathbf{E}_{AC}$ = load-winding voltage vector
e_{AC} = instantaneous value of load-winding voltage
E_L = load voltage, rectified average
$\mathbf{E}_L$ = load voltage vector
e_L = instantaneous value of load voltage
I_{DC} = average control current
i_{DC} = instantaneous value of control current
I_L = load current, rectified average
i_L = instantaneous value of load current
R_L = load resistance
C_L = capacitance paralleling the load resistance
θ = carrier phase angle between load-winding voltage and load voltage
ϕ = signal phase shift between control voltage and load voltage

Vectors are shown in the illustrations by a bar over the letters.

Theory

The average flux density (B_0) in each core of a magnetic amplifier depends not only on the magnetic intensity (H_0) applied to the core but also on the amplitude (B_m) of any superimposed alternating flux density.[1,2] This dependence on B_m in a single core is illustrated in Fig. 1 for the case of materials with gradually varying **B-H** curves. Points on the sine wave of flux-density variation are projected horizontally to the **B-H** curve and then down to the corresponding time on a new time axis. The average value of the resulting curve of H over one cycle is H_0 for the given values of B_0 and B_m. This H_0 is greater than the H corresponding to the given B_0 when B_m is zero.

Expressed another way, for a given average (d-c) ampere-turns on a core (or a pair of such cores), the average value of flux linking any winding on the core is reduced if an alternating voltage is applied to a winding on the core. (The average current in a winding is fixed by the average voltage and the resistance in the winding circuit. The core cannot affect this average current.) The ability to change the average flux linkage by varying the applied alternating voltage is the tool to be used in making the control winding of a magnetic amplifier noninductive.

The quantitative relationships between B_m, H_0, B_0, and load current (I_L) have been presented by others.[3,4,5] For reference, an oscillographic example is shown in Fig. 2 taken from the magnetic amplifier described in the Appendix. Here two families of curves on the same set of co-ordinates relate the four quantities.

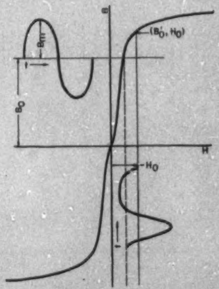


Fig. 1. Typical B-H relationship of magnetic material showing effect of an alternating flux density (B_m) on average flux density (B_0' without B_m, B_0 with B_m) for a fixed value of average magnetic intensity (H_0)

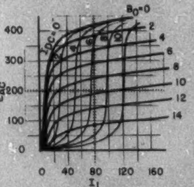


Fig. 2. Magnetic-amplifier output characteristics with superimposed lines of constant average flux density (B_0, in kilogauss). E_{AC} is the Rms value of voltage across the two load windings; I_L is the rectified average value of load current, milliamperes; and I_{DC} is the control current, milliamperes

---

Paper 58-774, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 1, 1958.

D. P. Chandler and R. W. Downing are with North American Aviation, Inc., Downey, Calif.

Vol.
Pt. 1
1958

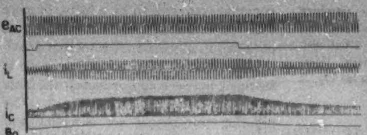


Fig. 3. Transient response of 400 cps with resistive load and conventional operation. e_{AC} is the voltage across the two load windings, i_L is the load current, i_C is the control current, and B_0 is the flux density averaged over several cycles. The extra line indicates initiation of the transient

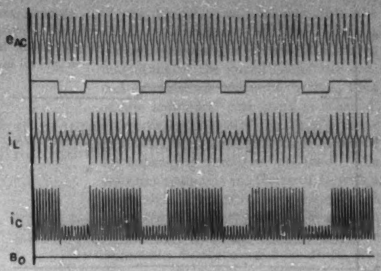


Fig. 4 (right). Transient response of 400 cps with resistive load switched synchronously with the signal in such a way as to keep B_0 constant. Symbols are same as for Fig. 3

For convenience, the alternating voltage (E_{AC}) across the load windings and the average value of control current (I_{DC}) have been used instead of B_m and H_0 to which they are respectively proportional. Note that for the major portion of the practical operating range, movement is possible along a constant B_0 curve from one set of control and load currents to another set without materially deviating from the turns-ratio law.[1,6] This law states that control ampere-turns nearly equal the load-winding rectified average ampere-turns.

To remain on a constant B_0 line it is necessary only that E_{AC} increase properly with increase load current. If this relationship is properly maintained a signal change will result in an instantaneous change in the amplitude of the load current. This is exactly true if the change occurs at an instant when the alternating flux density is zero. If the change occurs at any other time in the cycle, a slight shifting of flux densities between the cores is required, but as will be seen in the experimental results this shift takes place in an undetectably short period of time.

If E_{AC} is made to increase more rapidly, than just enough to keep B_0 constant, with increased load current, the control winding acts a negative inductance. This provides both the potentiality of derivative effects and the danger of instability.

Experimental Verification

To check the possibility of obtaining instantaneous response by making the magnetic-amplifier control winding noninductive, tests have been performed on a simple series magnetic amplifier (more completely described in the Appendix) using hypersil toroidal cores. This amplifier and its control circuit were purposely chosen to have a long transient time with a resistive load so as to make pronounced improvements possible. For all the tests, the control winding was driven from a low-impedance source, making the control winding the only effective resistance in the control circuit.

The transient response of this amplifier with resistive load and 400-cps (cycle per second) power is shown in Fig. 3. The effective time constant for these transients is approximately that which published equations would predict.[7,8]

In contrast, Fig. 4 is an oscillogram of the transient response of the same amplifier with part of the load resistance short-circuited at the same time the control voltage is increased (see the Appendix). By this means E_{AC} is increased the appropriate amount for the increase in I_L and I_{DC} so that the two steady-state conditions involved the same B_0. Admittedly this method is not practical for magnetic-amplifier applications, but it demonstrates the feasibility of making the control winding noninductive.

In order that the noninductive transient be seen in more detail, it was redone using 60-cps power; see Fig. 5. Here the instantaneous nature of the change can be seen. Transients initiated at four different phase angles of the supply power are shown in the oscillogram. Notice that the second transient occurred at an instant of zero load current, resulting in no jump, merely an instantaneous shift to a different but tangent waveform, as predicted by theory. The other three transients are more dramatic, however, actually demonstrating an instantaneous change to the new steady-state output. Superposition of portions of the two steady-state waveshapes onto the transient portions of the oscillogram indicates no detectable deviation from steady state on either side of any of the transient steps in the load current.

For still better detail, the third transient of Fig. 5 is enlarged as Fig. 6. Notice that the slope of the transient step in i_L is about the same as that for the other steps and is caused primarily by the response of the oscillograph galvanometer.

Fig. 7 is an oscillogram of the tran-

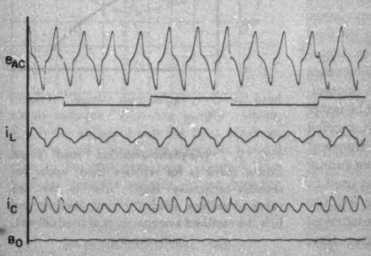


Fig. 5 (left). Transient response of 60 cps with resistive load switched synchronously with the signal in such a way as to keep B_0 constant. Symbols are same as for Fig. 3

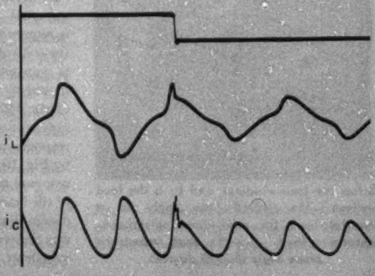


Fig. 6 (right). Enlargement of the third transient in Fig. 5

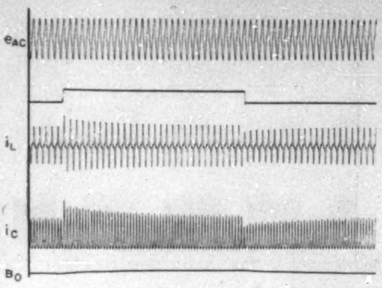


Fig. 7. Frequency response of 400 cps with resistive load switched as for Fig. 4 but with constant control voltage. Symbols are the same as for Fig. 3

sients resulting from switching the load circuit as for Fig. 4 but without switching the signal. Here the derivative capabilities of a negative control inductance is demonstrated.

Practical Application

One practical method of increasing E_{AC} with increased load current to improve the transient response is to make the load somewhat capacitive.

To get a "feel" for the effect of capacitance across the load, assume for the moment that the load winding of the magnetic amplifier is a variable linear inductance. Then with sinusoidal supply voltage and a linear load all the voltages appearing in the load circuit are sinusoidal and may be represented by vectors. Fig. 8 is a vector diagram of such voltages for the case of a purely resistive load (dashed lines) and for a resistive-capacitive load. The supply voltage E is the same for both cases. With the resistive load, there is always a 90-degree angle between the load voltage vector $\mathbf{E}_L$ and the vector $\mathbf{E}_{AC}$ of voltage across the magnetic-amplifier load winding. This results in a semicircle as the locus of the intersection of $\mathbf{E}_L$ and $\mathbf{E}_{AC}$ for various values of the load-winding inductance (controlled by changing the magnetic - amplifier control current). When the load is made capacitive the phase angle (θ) between the inductive load winding and the load impedance is greater than 90 degrees resulting in a voltage vector diagram as shown by solid vectors. The locus of the intersection of the $\mathbf{E}_{AC}$ and $\mathbf{E}_L$ vectors is now a portion of a circle such that $\mathbf{E}$ is a chord of the circle. An examination of these diagrams shows the capability of causing E_{AC} to increase (at least over a limited range) with increase of load current by an amount determined by how capacitive the load is made.

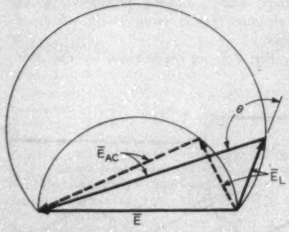


Fig. 8. Vector diagrams of load circuit voltages. E is the supply voltage, E_{AC} the voltage across the load windings, and E_L is the load voltage. The dashed lines apply for a resistive load, the solid lines for resistive-capacitive load. θ is the load impedance phase angle plus 90 degrees

For the authors' purposes, a more useful way of presenting this information is in the form of load lines[5] on the same set of co-ordinates used in Fig. 2. Fig. 9 is an oscillographic presentation of such load lines for a resistive load (lower curve) and the same load resistance paralleled by a capacitance. The capacitive load is seen to cause E_{AC} to increase with increasing load current even with the nonsinusoidal waveforms caused by the nonlinear magnetic amplifier.

By comparing Fig. 9 with Fig. 2, it is apparent that much less change in B_0 is required for a given load current change (or control current change) when the load is capacitively tuned. This is tantamount to saying the effective inductance of the control winding has been appreciably reduced. The result should be faster response to a transient. This is verified by Fig. 10, an oscillogram of the transient response resulting from capacitive tuning of the load in the amplifier circuit used to obtain Fig. 3. The tuning for Fig. 10 was limited to a value which provides no overshoot. Even so, the effective time constant was reduced by a factor of more than three. Still faster response can be obtained if an overshoot is tolerated.

The rectified average load current is swinging between approximately the same values for both Fig. 10 and Fig. 3. The differing waveshapes are deceiving and make the determination of effective time constants difficult.

Unfortunately the load lines obtainable by capacitive tuning of the load do not match exactly the constant B_0 lines for any significant variation in load current. So, from a practical standpoint, this method allows reduction of effective control-winding inductance but not elimination of it.

Application to Push-Pull Circuitry

Many applications such as driving servomotors require reversible polarity outputs from the driving amplifier. A push-pull arrangement (described in the Appendix) of two simple series-connected magnetic amplifiers is one way of fulfilling this requirement and presents some special problems worthy of consideration.

REVIEW

A review of steady-state conditions in such an amplifier is probably desirable. Because current is inductively limited in simple series magnetic amplifiers[5,6,9] it is possible to bias (by currents in control windings) both core pairs into a state of permitting about half the maximum desired load-winding current to circulate through the magnetic-amplifier load windings with no current going through the load. Then if the control currents are

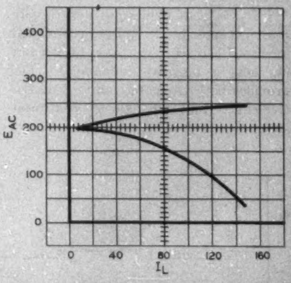


Fig. 9. Magnetic-amplifier load lines. Lower curve is for resistive load, upper for resistive-capacitive load. E_{AC} is the rms value of voltage across the load windings and I_L is the rectified average value of load current, milliamperes

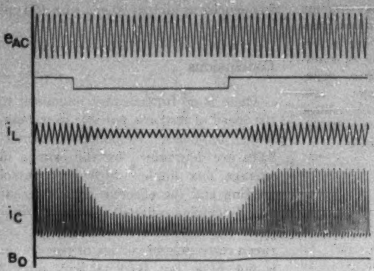


Fig. 10. Frequency response of 400 cps with resistive load tuned by parallel capacitor. Symbols are same as for Fig. 3

unbalanced, the load-winding currents are no longer balanced and their difference must flow through the load. If the control currents are unbalanced in the opposite direction, the polarity of the current through the load is reversed. Because in each of the core pairs the rectified average load circuit ampere-turns are nearly equal to the control ampere-turns,[1,4] the load current is approximately proportional to the difference in control currents and is nearly independent of the sum of the two control currents within the bounds of the operating range of the amplifier.

When a resistive load and a core material with gradually varying **B-H** curve are used, the load current and voltage are roughly 90 degrees out of phase with the supply voltage, and do not require phase shifting of power for operation of 2-phase servomotors. Fig. 11(A) is a vector diagram of the voltages in the load circuit for a given output into a resistive load. The solid vectors represent voltages when core pair 1 (whose voltage vectors carry the subscript 1) is conducting more than pair 2. The dashed vectors apply when the polarity of the output is reversed, with core pair 2 conducting more than pair 1. The curved line is an approximate locus of the end of the load voltage vector for various outputs. Notice that at all times the vector sum $(\mathbf{E}_{AC_1}+\mathbf{E}_{AC_2})$ of the voltages across the two pairs of load windings must equal the fixed total supply voltage $(\mathbf{E}_1+\mathbf{E}_2)$ regardless of the load voltage. Among other things this means that the sum of the voltage amplitudes on the load windings must increase as the load voltage increases from zero in any direction other than being in phase or 180 degrees out of phase with the supply voltage

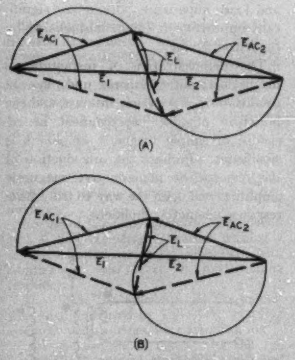


Fig. 11. Vector diagrams of push-pull load circuit voltages. $\bar{E}$ is the supply voltage, $\bar{E}_{AC}$ is the voltage across the load windings, $\bar{E}_L$ is the load voltage, and subscripts 1 and 2 refer to the two core pairs. Dashed lines hold for opposite polarity output from the solid lines

A—Resistive load
B—Resistive-capacitive load

EFFECTS OF CAPACITIVE TUNING

Because the load current is nearly 90 degrees out of phase with the voltage across the core pair supplying the load current, the phase angle of the load voltage is controlled by the phase angle of the load impedance. In particular, making the load capacitive tilts the load voltage vector in the direction to increase the voltage across the core pair conducting the greater load circuit current and reducing the voltage across the other core pair. This effect is in the proper direction for reducing the change in B_0 from one operating point to another, so again the capacitive load is tending to improve the transient and frequency response of the amplifier. Fig. 11(B) is a voltage diagram and approximate load vector locus for a capacitively tuned load.

Load-line considerations of the push-pull circuit are complicated by the possibility of shifting the bias on both halves of the amplifier without affecting the output significantly. If the bias is changed as a function of the instantaneous magnitude of the signal (difference control current), the shape of the load lines of the individual core pairs is affected. However, regardless of the shape of these load lines, their average slope is increased by capacitive tuning of the load. The resulting reduction in change of B_0 with a given signal change reduces the effective control-winding inductance. Too much capacitance can result in a negative inductance in the control windings, producing an oscillatory condition.

In practice nearly the same improvement in response can be obtained by load tuning in the push-pull circuit as for the single core pair. The result is a practical, simple magnetic amplifier suitable for driving servomotors, especially in torque motor applications in which the motor speed is limited, preventing large changes in the motor impedance.

Frequency Response

The frequency response of an amplifier determines the effect it has on the stability of closed-loop servomechanisms. The only precaution necessary for predicting stability, even with nonlinear amplifiers, is that the signal phase shift and gain as functions of frequency must be obtained for the same amplifier conditions which exist in the servomechanism. It is therefore desirable to consider frequency-response capabilities of the noninductive magnetic amplifier.

The use of a signal-modulated carrier frequency does not necessarily imply a

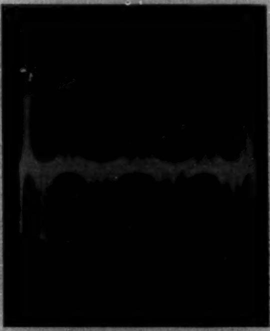


Fig. 12. Lissajous pattern with load current as ordinate and control voltage as abscissa. Supply frequency is 400 cps, signal frequency about 40 cps

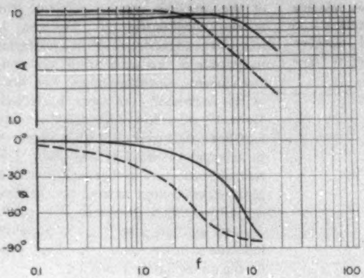


Fig. 13. Frequency response of single-polarity-output magnetic amplifier. Supply frequency is 400 cps. A is voltage gain, ϕ the phase angle, and f the frequency, cps. Dashed curves are for resistive load, solid curves for resistive-capacitive load

time delay or a phase lag as is often assumed. If the envelope of an amplitude-modulated signal at the output of an amplifier is in phase with the signal input, the amplifier does not contribute to system instability at that frequency. Some magnetic amplifiers, such as the Ramey circuit,[10,11] do provide a half-cycle delay resulting in a signal phase shift proportional to frequency. This delay is not inherent in all magnetic amplifiers, however.

It is necessary, of course, to avoid signal frequencies so high that the modulation is meaningless.

A magnetic amplifier with a truly noninductive control winding produces no phase lag. By keeping the signal amplitude small and carefully adjusting the load tuning capacitance for a given frequency so as to make the load line match a constant B_1 line for that quite limited operating point, zero phase shift can be produced at frequencies as high as one tenth of the supply frequency. Fig. 12 is a Lissajous pattern of output versus input signals showing the result of such adjustment of the magnetic amplifier described previously. The signal frequency is about 40 cps with a 400-cps power frequency. Without the tuning capacitance, approximately 90-degree lag results. A slight increase in capacitance makes the amplifier oscillatory.

For practical operation over reasonable signal ranges it is necessary to reduce the load tuning capacitance somewhat in order to avoid amplifier instability. The resultant slightly inductive magnetic amplifier does provide phase lag but is still a considerable improvement over the same amplifier with resistive load. Fig. 13 compares the frequency response of such an amplifier tuned to a practical degree and with a purely resistive load.

The frequency response of a pair of amplifiers connected in push-pull, driving a locked-rotor servomotor as the load, is shown in Fig. 14 for the cases of the load being tuned for unity power factor and (with increased capacitance) for improved frequency response. As with the single-ended magnetic amplifier, the load tuning has been limited to a value which retains good signal waveform. Higher frequency response is possible with degradation in waveform. It should be borne in mind that driving this amplifier from a high impedance source as would typically be done improves the response still further. The result is an amplifier suitable for use in servomechanisms requiring negligible phase lag at signal frequencies as high as 20 cps with 400 cps power.

Conclusions

There is no fundamental limitation to the speed of response possible from magnetic amplifiers. Response characteristics are determined by the change in average flux linkage with the control winding and the effective control resistance (including reflected resistances from other windings). Any relationships between response and voltage or power gain is not basic but is dependent, among other things, on the type of load used.

When no average flux change occurs during a transient, windings inductively coupled to the control winding are of no consequence. Even eddy currents and field propagation delays in the cores cannot cause delays when no change in flux is required.

One method of providing noninductive control is capacitive tuning of the loads of magnetic amplifiers in which the load current is inductively limited. Indeed, too much capacitance across the load can make a negative inductance of the control winding. The instability resulting from a negative inductance imposes the practical limitations of requiring a fixed load impedance and less than perfect tuning to allow for deviations between the shapes of the load line and the constant average flux density line, and for variations in supply voltage, supply frequency, bias, and load impedance. Even so, significant improvement can be had reliably by capacitive tuning.

The demonstration of instantaneous magnetic-amplifier response under special conditions is of academic interest, and the presently possible improvement in response of simple magnetic amplifiers is significant. Perhaps the introduction of the concept of noninductive magnetic amplifiers will open the way to truly fast-response magnetic amplifiers.

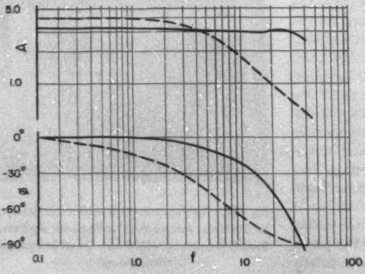


Fig. 14 (left). Frequency response of push-pull magnetic amplifier driving locked-rotor 400-cps servomotor. A is voltage gain, ϕ the phase angle, and f the frequency, cps. Dashed curves are for resistive load, solid curves for resistive-capacitive load

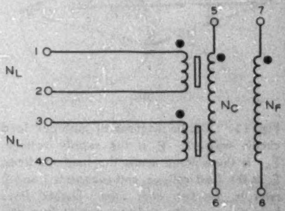


Fig. 15. Schematic of saturable reactor used for tests: N_L=1,000 turns, N_C=12,000 turns, N_F=500 turns

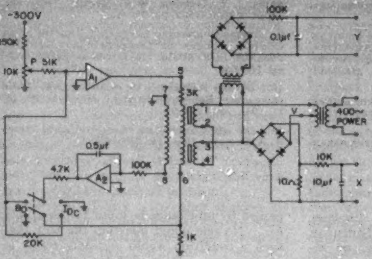


Fig. 16. Schematic of circuit used to produce Fig. 2. A_1 and A_2 are operational amplifiers; X and Y go to corresponding inputs of an oscilloscope

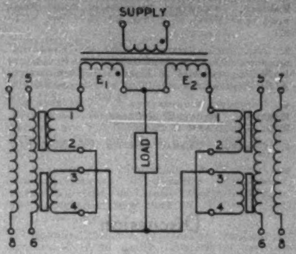


Fig. 18. Push-pull reversible output a-c magnetic amplifier

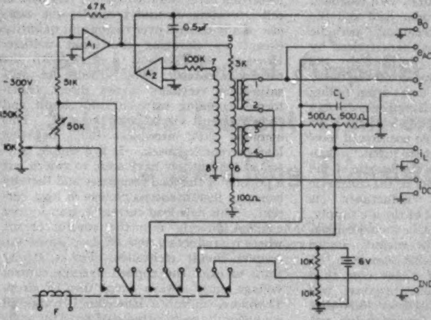


Fig. 17. Schematic of circuit used to obtain transient response. A_1 and A_2 are operational amplifiers. F is supplied from a square-wave function generator, and all output terminals go to the oscillograph used to photograph the transients. IND is the signal indicating initiation of the transient

Appendix. Experimental Equipment

The saturable-reactor core pair used for all the experimental work described is shown schematically in Fig. 15. Each of the cores are 4-mil tape-wound hypersil toroids of 1.5-inch outside diameter, 1.0-inch inside diameter, and 0.375-inch height. A 1,000-turn load winding (N_L) is wound on each core. Then the 3,000-ohm 12,000-turn control winding (N_C) is wound on the pair of cores stacked together. The 500-turn feedback pickoff winding (N_F) is added for sensing changes in control flux linkages.

For all of the tests described terminals *2* and *4* were connected together making a series-connected magnetic amplifier, with terminals *1* and *3* as its load-winding terminals.

Fig. 16 is the schematic of the circuit suggested and used by John Cox, research engineer at Autonetics, for generating Fig. 2. When the switch is thrown to the I_{DC} position, the control current is controlled to the value determined by the setting of potentiometer *P* regardless of changes in a-c supply voltage. For each value of control current one of the constant I_{DC} lines is traced out as the alternating voltage is varied by turning the Variac *V*. When the switch is thrown to B_0, the control current is controlled in such a way as to maintain B_0 constant. Curves of constant B_0 can thus be traced by varying the alternating voltage for various values of B_0. By photographing the moving spot on the oscilloscope for several constant B_0 and several constant I_{DC} curves, Fig. 2 was obtained.

The load lines of Fig. 9 were obtained in a similar manner. A load of a 1,000-ohm resistor or this resistor paralleled by a 0.53-microfarad capacitor were added in series in the load circuit. With the Variac fixed at the proper position and the switch in the I_{DC} position, the load lines were traced out by varying the control current.

The transient-response oscillograms were taken using the circuit of Fig. 17 or modifications of it. The circuit as shown with different load resistors for Fig. 4 and with C_L being zero was used to obtain the instantaneous response of Figs. 4 through 6. By disconnecting the appropriate one of the relay switching circuits and adding capacitance to the load when needed, each of the other transients shown were obtained.

It should be noted the 3,000-ohm winding resistance in the control winding is the total effective control circuit resistance because of the very low output impedance of closed-loop operational amplifiers.

Frequency-response data were obtained by introducing a sinusoidal signal into the operational amplifier input and comparing the output of the magnetic amplifier with this input. Again, the only control circuit resistance is the 3,000-ohm winding resistance.

The push-pull circuit used for reversible output was made up from two of the saturable reactors shown in Fig. 15. Their load windings were connected as shown in Fig. 18. For the frequency-response tests, the control windings were driven in push-pull by two separate operational amplifiers, each connected the same as the one used for driving the single core pair.

Values of the variables associated with the transient- and frequency-response tests are given in Table I.

References

1. MAGNETIC AMPLIFIERS, A. G. Milnes. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. I, May 1949, pp. 89–98.

2. AN ANALYSIS OF INTERLINKED ELECTRIC AND MAGNETIC NETWORKS WITH APPLICATIONS TO MAGNETIC AMPLIFIERS, D. W. Ver Planck, M.

Table I. Data for Transient- and Frequency-Response Tests

Fig. No.	Drive Frequency, CPS	R_L, Ohms	C_L, Microfarad	Average Flux Change, Maxwells	E_{AC}, Volts, RMS	I_{DC}, Milliamperes	I_L, Milliamperes	Power Frequency, CPS
3	3	1,000	0	4,500	250 to 195	2.0 to 6.2	25 to 70	400
4	29	5,400 to 1,000	0	0	235 to 255	2.0 to 6.2	25 to 72	400
5 and 6	10	1,000 to 500	0	0	30 to 37	2.8 to 3.7	30 to 47	60
7	10	1,000 to 500	0	−400	250 to 265	4.0	50 to 55	400
10	7	1,000	0.56	900	250 to 265	2.0 to 6.2	25 to 73	400
12	40	1,000	0.563		255	6.3	75	400
13	0.1 to 40	1,000	0		212 to 260	3.3 to 5.3	12 to 85	400
	0.1 to 40	1,000	0.39		210 to 220	3.2 to 5.2	20 to 82	400
14	0.1 to 40	*	0.59			1.4 to 7.0	15 to 95	400
	0.1 to 40	*	0.68			2.4 to 5.2	20 to 105	400

* Kearfott *R111-2* servomotor; impedance = $640/\underline{64°}$ ohms.

Fishman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, 1949, pp. 1012–27.

3. ANALYTICAL DETERMINATION OF CHARACTERISTICS OF MAGNETIC AMPLIFIERS WITH FEEDBACK, **D. W. Ver Planck, L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 565–70.

4. UNIVERSAL CURVES FOR D-C CONTROLLABLE REACTORS, **Walter C. Johnson, B. C. Merrell, R. E. Alley, Jr.** *Ibid.*, pp. 31–40.

5. GENERAL CHARACTERISTICS OF MAGNETIC AMPLIFIERS, **L. A. Finzi, D. C. Beaumariage.** *Ibid.*, vol. 69, pt. II, 1950, pp. 919–27.

6. A THEORETICAL AND EXPERIMENTAL STUDY OF THE SERIES-CONNECTED MAGNETIC AMPLIFIER, **H. M. Gale, P. D. Atkinson.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. I, May 1949, pp. 99–114.

7. RESPONSE TIME OF MAGNETIC AMPLIFIERS, **E. L. Harder, W. F. Horton.** *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1130–41.

8. THE FIGURE OF MERIT OF MAGNETIC AMPLIFIERS, **J. T. Carleton, W. F. Horton.** *Ibid.*, vol. 71, pt. 1, Sept. 1952, pp. 239–45.

9. SERIES-CONNECTED SATURABLE REACTOR WITH CONTROL SOURCE OF COMPARATIVELY LOW IMPEDANCE, **H. F. Storm.** *Ibid.*, vol. 69, pt. II, 1950, pp. 756–65.

10. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, **Robert A. Ramey.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

11. ON THE CONTROL OF MAGNETIC AMPLIFIERS, **R. A. Ramey.** *Ibid.*, pp. 2124–28.

◆

Discussion

R. M. Hubbard (Boeing Airplane Company, Seattle, Wash.): The authors have succeeded in presenting plausible explanations for two different phenomena which they attribute to the elimination of the control circuit inductance of a saturable-reactor type of magnetic amplifier. While one cannot dispute the fact that the experimental evidence does indeed indicate that fast response has been achieved, it is felt that the authors' historically acceptable analysis technique is not entirely convincing. Therefore an effort will be made to suggest an alternate means of explaining the observed phenomena.

In this discusser's opinion, the magnetic-amplifier art has reached a point in its development at which it is no longer necessary to use the analytical crutch of referring to the magnetic-amplifier control circuit as being characterized by a "sort of inductance." By denying the existence of such an inductance, one finds it relatively easy to explain its apparent elimination. Of course, the very existence of windings and, under some conditions, of an exponential-like response clearly suggests an inductive control circuit. Unfortunately, such an assumption puts one in the position of having to justify an inductance which, rather than having a magnitude dependent on the number of turns on a magnetic core, actually has its magnitude determined by the turns ratio, the value of the load resistance, and even the a-c supply frequency. Further, why does the control winding on a single-core amplifier exhibit no apparent inductance, while a combination of two single-core amplifiers is characterized by a highly inductive control circuit? How does one explain the fact that decreasing the L-R time constant by increasing the control circuit resistance yields radically different results between the case of the self-saturating circuit and that of the saturable-reactor circuit? In particular, the response of the former circuit approaches a pure delay while that of the later circuit approaches "instantaneous response," and neither circuit is characterized by exponential response. This discusser would like to suggest that the answers to these questions are beclouded by the practice of associating an equivalent linear inductance with magnetic-amplifier control circuits. Over restricted operating ranges, such an expediency yields usable results but when one is interested in the regions of fast response, the linear analogy crutch fails completely. It is further suggested that the answers to the subject questions can readily be found by simply recognizing the fact that the magnetic amplifier is a feedback system incorporating a time delay. This discusser has shown that such systems readily display exponential response, pure delay, or instantaneous response without the necessity for introducing the concept of inductance.[1,2] In fact, the use of delay lines with feedback is growing in both the radar field and the controls field, and the resultant "periodic filters" are in many cases characterized by exponential response.[3,4] Fortunately the absence of windings in many of the delay lines has prevented the users from falling into the trap of characterizing these devices by inductive input circuits. They have thus avoided the untenable position of having to explain a device whose input "inductance" is proportional to the length of the line in a manner analogous to the magnetic amplifier whose input "inductance" is proportional to the period of the a-c supply.

The authors are correct in their observation that if a change in output can be achieved with no appreciable change in the average flux linkages of either core then substantially instantaneous response will result. However, this discusser feels that the achievement of this condition has not been clarified by the authors' description of the theory. For example, can the technique for achieving instantaneous response by switching the load simultaneously with the input be achieved in the self-saturating class of circuit? The answer is obviously, "No, switching the load resistance can have no effect on transient response." Let us examine the reason for this basic difference. If the rectifiers are well behaved, any manipulation of the load circuit of a self-saturating circuit can only affect the gating core. Clearly the value of the average flux linkages of the gating core is predetermined and it is impossible to change the half-cycle average voltage across its load winding either by switching the load or by increasing the supply voltage. Thus E_{AC} cannot be increased with increasing load current and no modification of transient response is possible.

The important conclusion that should be drawn from the foregoing observations is that, in order to change E_{AC}, one must change the average flux linkages of the resetting core. This is true in both the self-saturating and in the saturable-reactor circuits and, in fact, any change in E_{AC} can be the result of only a change in the average load-winding voltage of a resetting core. Now the problem of achieving instantaneous response can be restated. How can one achieve a steady-state change in load current, and at the same time apply a compensating change to the average volts per turn of the windings on the resetting core so as to maintain constant the average volts per turn of that core? In essence one is asking that a change in output be accompanied by no change in input to the amplifier. No one could question the fact that such a condition would certainly result in no transient. It is obvious that the decoupling between load circuit and input (resetting core) afforded by the rectifiers in self-saturating circuits precludes the exercise of any control over the output quantity except by changing the input voltage. However, such is not the case with the saturable reactor. The absence of self-saturating rectifiers allows direct transformer coupling between load circuit and control circuit via the resetting core. This coupling path introduces 100% negative load current feedback. It is a simple matter, as the authors have seen, to switch out a portion of the load resistance and thereby cause an instantaneous change in load current. This new load current is transformer coupled directly into the control circuit where it produces a new IR drop across the control circuit resistance. For a steady state to exist the applied average control voltage must always equal this IR drop. Therefore, unless a cancelling change in control voltage is introduced an error voltage will be applied to the input (resetting core). If such a voltage is applied as input to the resetting core, our efforts to achieve a noninductive transient will not have been realized. Instead, a transient will ensue which will finally return the load current to its original value even though the load resistance has been changed. Once the original value of current is established, the resultant IR drop in the control circuit will again equal the constant applied control voltage and a steady state will exist. This so-called derivative effect, as described by the authors, is merely a demonstration of the fact that in a high-gain amplifier with 100% load current feedback, changes in load resistance can have only transitory effects upon the load current. It should be clear at this point that, if the load resistance and the control voltage are switched simultaneously, the change in control voltage can exactly compensate the change in control circuit IR drop caused by the load current. Thus no input would be applied to the resetting core. The net result is a change in output with no change in input, or a noninductive transient. One, of course, can question that the "brute force" technique of actually switching the output to achieve an instantaneous change in output can properly be termed "elimination of control circuit inductance." Also, it should be obvious that there is no danger of instability associated with the derivative effects discussed in the foregoing.

It is felt that the effect of a capacitive load on transient response involves a different phenomenon than is illustrated by the load-switching experiment. It is the discusser's

Vol.
Pt. I
195

opinion that the addition of capacitance to the load introduces a feedback phenomenon due to the resultant overlap or commutation effects. It has been shown by others that feedback can indeed produce fast, but not instantaneous, response as well as lead to stability problems in magnetic amplifiers. The fact that, unlike the previous case, the addition of capacitance to the load can produce equally good improvements in the transient response of both self-saturating and saturable-reactor classes of magnetic amplifiers suggests a somewhat different approach to the analysis problem.

It was previously stated that changes in the load circuit could not affect the resetting core in a self-saturating circuit and therefore could have no effect on transient response. However, the qualification that the rectifiers be well behaved is not actually satisfied in a capacitive circuit. The load voltage actually overlaps between the gating and the resetting half-cycles so that the load voltage does indeed effect the voltage applied to the resetting core. This constitutes negative feedback since an increasing load voltage applies an increasing voltage to the resetting core. The same sort of phenomenon occurs in the saturable-reactor circuit, and the net result is that transient response is speeded up by the negative feedback in the case of either circuit. The use of too much capacitance can result in the introduction of excessive negative feedback which produces an oscillatory condition. Thus, since control circuit inductance is nothing more than a manifestation of intrinsic feedback produced by the circuit connection, it is not surprising that the introduction of feedback can in fact "eliminate the inductance."

REFERENCES

1. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, Robert M. Hubbard. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

2. BLOCK-DIAGRAM METHOD OF ANALYSIS APPLIED TO THE SATURABLE REACTOR, R. M. Hubbard. *Ibid.*, vol. 77, pt. I, Mar. 1958, pp. 57–65.

3. ANALYSIS AND SYNTHESIS OF DELAY LINE PERIODIC FILTERS, H. Urkowitz. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-4, no. 2, June 1957.

4. POSICAST CONTROL OF DAMPED OSCILLATORY SYSTEMS, O. J. M. Smith. *Proceedings*, Institute of Radio Engineers, vol. 45, no. 9, Sept. 1957, pp. 1249–55.

A New Control Amplifier Using a Saturable Current Transformer and a Switching Transistor

RAY E. MORGAN
MEMBER AIEE

THE PERFORMANCE of a large portion of electric equipment relies on dependable control amplifiers. Many control systems require fast-responding, compact, and economic amplifiers along with the demand for reliability. Saturable magnetic circuits combined with switching transistors promise good solutions for these control requirements.

This paper describes a magnetic-transistor amplifier that uses a saturable current transformer and a switching transistor; see Fig. 1. A 1-watt transformer and a transistor dissipating 2 watts proportionally control a 300-watt load. The saturable transformer transistor circuit is an oscillator, shown in Fig. 2, which is similar to the oscillator described in reference 1, except that a saturable current transformer is used in place of the saturable voltage transformer used in reference 1, and in Fig. 2 the load is placed in series with the transformer and transistor. The frequency of the oscillator of Fig. 2 is controlled by a direct current in a winding of the saturable transformer similar to the frequency controlled oscillator of reference 2, except in the circuit of Fig. 2, the load current I_L is controlled, as shown in the transfer characteristic of Fig. 3.

This paper describes methods of controlling the magnetic-transistor amplifier by a direct current and a variable resistor, and by a small integral magnetic amplifier that receives its power from the saturable current transformer transistor oscillator. The Delco *2N174* transistor is used, for the data in this paper and the saturable transformer core is made of 79% nickel, 17% iron, and 4% molybdenum.

Principle of Control

Basically, the magnetic-transistor amplifier operates as a chopper similar to the operation of the switch chopper of Fig. 4(A). The load current i_L waveshapes are shown in Fig. 4(B), (C) and (D), which are provided by opening and closing switch S_1, and these load current wave shapes are the same as those provided by the magnetic-transistor amplifier. The length of time that switch S_1 is closed remains approximately constant, while the length of time it is open is varied. The shorter the time it is open, the greater the load power becomes, while the longer the time it is open, the smaller the load power becomes. The load power of Fig. 4(A) is controlled by opening and closing switch S_1, which varies the time that the switch is open in relation to the time it is closed.

Operation of the Magnetic-Transistor Amplifier

The magnetic-transistor amplifier (Fig. 2) operates similarly to the switch chopper circuit of Fig. 4(A). The circuit of Fig. 2 oscillates and opens and closes the load circuit in the manner of switch S_1, Fig. 4(A), providing load current i_L waveshapes as shown in Fig. 4(B), (C) and (D).

Usually, the magnetic-transistor amplifier of Fig. 2 uses bias for the transistor, as shown in Fig. 5, with the use of resistors R_1 and R_2. The position of the transistor bias is shown in Fig. 6 (the steady-state characteristic of the transistor). The application of transistor bias

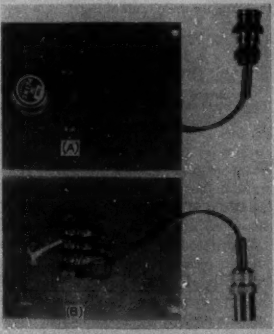


Fig. 1. Magnetic-transistor amplifier for controlling 300-watt load. A—Front view. B—Rear view

---

Paper 58-858, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 17, 1958; made available for printing April 18, 1958.

RAY E. MORGAN is with the General Electric Company, Schenectady, N. Y.

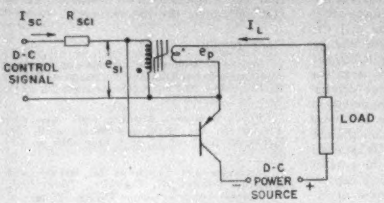
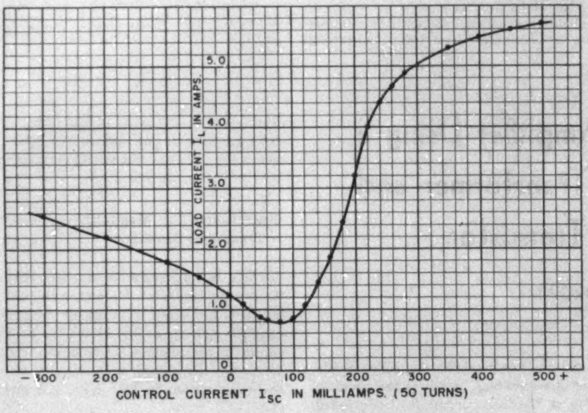


Fig. 2 (left). Circuit of magnetic transistor amplifier of Fig. 1, using saturable current transformer and switching transistor

Fig. 3 (below). Transfer characteristic of circuit of Fig. 2



permits the circuit (Fig. 5) to maintain oscillation with a wider variation in the power source voltage (5% to 100% of the transistor's rated voltage) and ambient temperature, than that of the circuit of Fig. 2 without transistor bias. The transfer characteristic of Fig. 3 (obtained by the circuit of Fig. 5) is approximately the same for the circuit of Fig. 2.

Consider the operation of the circuit of Fig. 5 with control current $I_{SC}=+300$ ma (milliamperes); see Fig. 3. The circuit oscillates with the saturable current transformer core tracing the **B-H** hysteresis loop of *abcda* of Fig. 7. The secondary voltage e_{S1} is shown in Fig. 8(A). Starting with the transformer core at negative saturation (at point *d* of Fig. 7), the application of transistor bias supplied by resistors R_1 and R_2 (Fig. 5) permits a small amount of load current i_L to flow through the primary of the current transformer (forcing the transformer core flux density to point *a* of Fig. 7), producing secondary voltage e_{S1} see Fig. 5. Secondary voltage e_{S1} adds to the transistor bias voltage in a regenerative manner and forces the transistor to "snap full on." The transistor remains full on during the flux density excursion from *a* to *b* of Fig. 7. At point *b*, the saturable current transformer saturates. The secondary voltage (Fig. 5) now drops to $e_{S1}=0$, and the transistor turns almost off, decreasing the load current i_L in the primary winding. As current i_L decreases, voltage e_{S1} is induced negative (negative at the dot in Fig. 5) forcing the transistor to snap off ($i_L \cong 0$). With the transistor turned off, load current i_L and the transistor base current are negligible, leaving current I_{SC} to control the transformer core. Current I_{SC} forces the transistor to be held OFF by negative voltage e_{S1} (the flux density dropping from *c* to *d* of Fig. 7) until the saturable current transformer core reaches negative saturation at point *d*. After negative saturation, the secondary voltage drops to $e_{S1}=0$ and the transistor bias partially turns on the transistor sufficiently for the regenerative action of the current transformer to force the transistor to snap on and the oscillation continues. Distributed capacitance in the current transformer winding helps the circuit to snap on and snap off in the manner described in reference 1 in the saturable voltage transformer oscillator.

The reduction of control current to $I_{SC}=+200$ ma (Fig. 3) changes the **B-H** hysteresis loop to *abefa* in Fig. 7 and the secondary voltage e_{S1} waveshape is shown

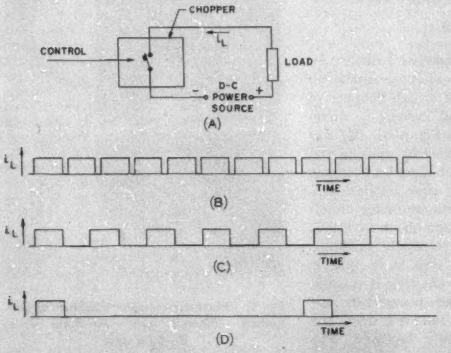
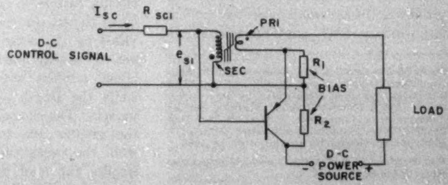


Fig. 4 (left). Switch chopper simulating operation of circuit of Fig. 2 and load current I_L waveshapes of Fig. 2. A—Switch simulator circuit. B—i_L at maximum output. C—i_L at one-half maximum output. D—i_L at minimum output

Fig. 5 (below). Same as Fig. 2 with transistor bias by resistors R_1 and R_2 added



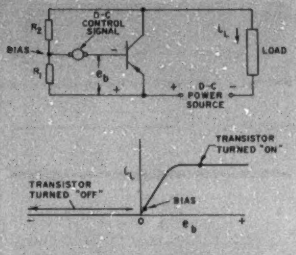


Fig. 6. Transistor circuit and steady-state characteristic of transistor, when not used as a switch, showing position of bias and position of characteristic of transistor when it is turned on and turned off

in Fig. 8(B). Likewise, reducing the control current to $I_{SC}=+100$ ma, the **B-H** loop is *abgha* (Fig. 7) and the waveshape of e_{S1} is shown in Fig. 8(C). In the same manner that high-frequency **B-H** hysteresis loops are wider (with a greater amplitude of H) than low-frequency loops, the greater the applied negative H (magnetomotive force in oersteds), the faster the flux density of the core traverses from positive to negative saturation. The time for the flux density excursion from c to d is less than the time for the flux density excursion from e to f and the time of the excursion from e to f is less than the excursion from g to h in Fig. 7. During the excursion of flux density from positive to negative saturation of the saturable current transformer (Fig. 5), control current I_{SC} is the main contributor of H to the core. With secondary voltage e_{S1} negative at the dot in Fig. 5 (forcing e_b of the transistor negative), the three terminals of the transistor are essentially open-circuit. For this condition, full control of the time the transistor is turned off is determined by current I_{SC}.

Decreasing control current to $I_{SC}<+100$ ma (Fig. 3) increases the frequency of the oscillator (Fig. 5) and gradually increases the load current I_L (Fig. 3). With $I_{SC}<+100$ ma, the saturable transformer traverses a minor **B-H** loop such as loop *nbmn* of Fig. 7. The flux density excursion of m to n being less than the excursion of g to h, the time the transistor is turned on and the time turned off is less for the **B-H** loop *nbmn* than **B-H** loop *habgh*, making the frequency increase as described in reference 2. As current I_{SC} is decreased below $I_{SC}=+100$ ma, the time off decreases more rapidly than the time on causing load current I_L to increase.

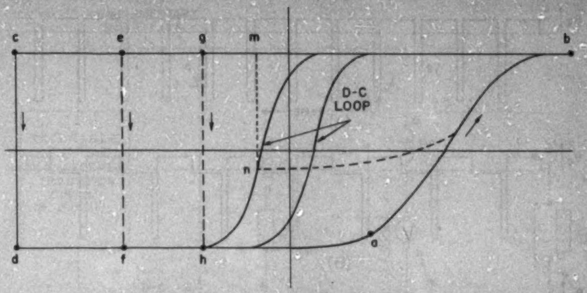


Fig. 7. B-H loops (flux density in kilogausses versus magnetizing force in oersteds) for saturable current transformer core at maximum load abcda, one-half maximum load abefa, and minimum load abgha

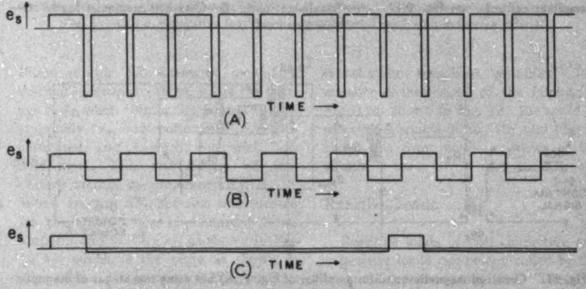


Fig. 8. Voltage waveshapes of saturable current transformer secondary voltage e_{S1} of Fig. 5. A—maximum load. B—One-half maximum load. C—minimum load

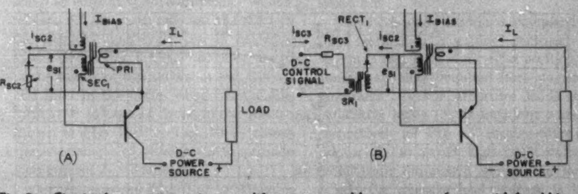


Fig. 9. Circuit of magnetic-transistor amplifier using saturable current transformer with fixed bias. A—Variable resistor control. B—Half-wave type of magnetic-amplifier control

The time the transistor is turned on remains constant as long as the saturable current transformer core maintains the full excursion from negative to positive saturation (minor **B-H** loops are avoided). During the time the transistor is turned on (a to b, Fig. 7), the saturable transformer operates as a regular current transformer. The power is supplied to the transformer by the primary current (i_L Fig. 5) and the secondary voltage is set by the secondary current and the secondary circuit resistance. The time the transistor is turned on is determined by $t_{ON}=N_{S1}\,\Delta\Phi/e_{S1}$, where N_{S1} is the number of turns on the secondary of the saturable transformer and $\Delta\Phi$ is the flux excursion of the transformer core. The turns N_{S1} and flux excursion $\Delta\Phi$ are design parameters of the saturable transformer. The secondary voltage e_{S1} is a function of design parameters of the saturable transformer, the secondary circuit components, the power source voltage, and load resistance.

For a given supply voltage, load, and

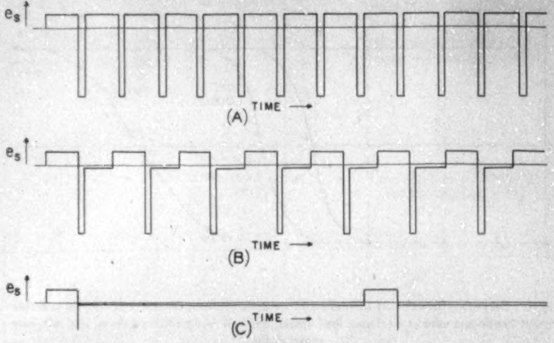


Fig. 10. Voltage waveshapes of saturable current transformer voltage e_{S1} with magnetic-amplifier control; see Fig. 9(B). A—Maximum load. B—One-half maximum load. C—Minimum load

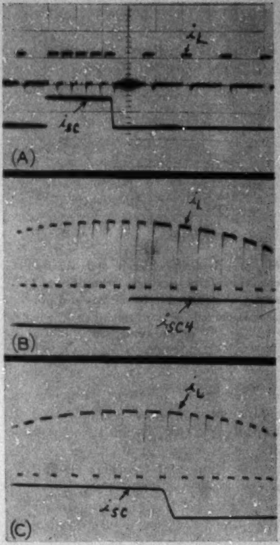


Fig. 13. Response time of (A) circuit of Fig. 5; (B and C) circuit of Fig. 11 using d-c power source from rectified unfiltered 60-cycle-per-second power. Oscillogram length, 5 milliseconds

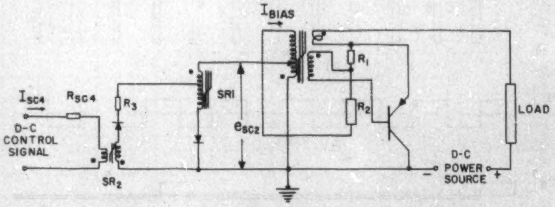


Fig. 11. Circuit of magnetic-transistor amplifier of Fig. 9(A) but using two stages of magnetic-amplifier control with first stage SR_2 controlling second stage SR_1 by (tapped) portion of gate-winding

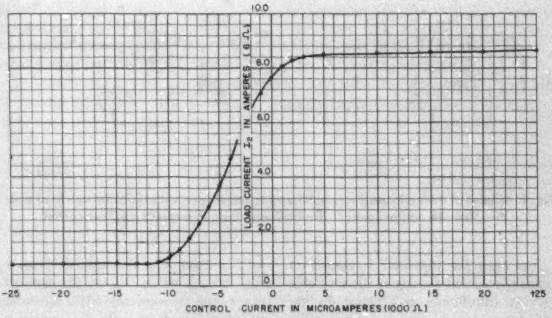


Fig. 12. Transfer characteristic of Fig. 11

design, the time the transistor is turned on remains constant.

The average load current is controlled by the control current I_{SC} varying the time the transistor of the oscillator circuit (Fig. 5) is turned off.

Variable Resistor Control

The saturable current transformer transistor amplifier can be controlled by a variable resistor such as a rheostat, a thermistor, etc. The amplifier circuit for resistor control is shown in Fig. 9(A) (transistor bias is omitted for simplicity). Fig. 9(A) is the same as Fig. 2 except that control current I_{SC} is replaced by a fixed bias current I_{BIAS} (separate winding), rectifier $RECT_1$, and a variable resistor R_{SC3}. The bias current I_{BIAS}, Fig. 9(A) is set to provide load current of $I_L=5$ amperes (Fig. 3) with resistor $R_{SC3}=\infty$; see Fig. 9(A). With resistor $R_{SC3}=\infty$, the waveshape of secondary voltage e_{S1} is shown in Fig. 8(A), which shows the same waveshape provided by the circuit of Fig. 2. As resistor R_{SC3}, Fig. 9(A), is decreased, secondary current I_{SC3} increases (when e_{S1} is negative at the dot and flux density is dropping between c and d of Fig. 7). Current i_{SC3} provides ampere turns in winding SEC_1 that oppose the ampere turns provided by the bias winding. As current i_{SC3} increases, load current I_L decreases in similar manner to the circuit of Fig. 2 when control current is decreased from $I_{SC}=+300$ ma in Fig. 3. Load current I_L decreases as control resistor R_{SC3} Fig. 9(A), decreases until $R_{SC3}=0$, and then the forward resistance of rectifier $RECT_1$

Vol
Pt.
195

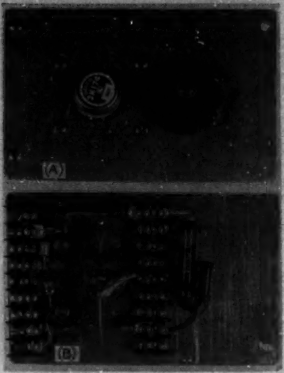


Fig. 14. Circuit of Fig. 11. A—Front view. B—Rear view

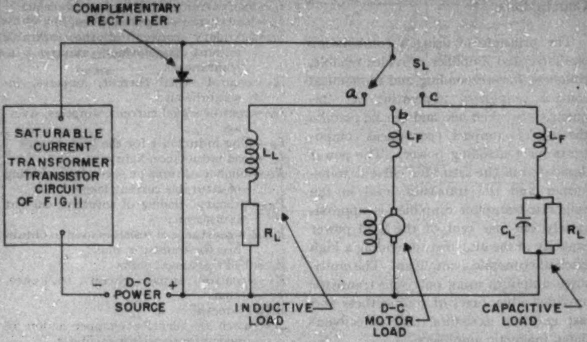


Fig. 15. Circuit of reactive loads for magnetic-transistor circuit of Fig. 11 using complementary rectifier with loads selected by switch S_L set at a, inductive; b, motor armature with inductance filter L_L, and c, capacitive load or capacitor-filtered load with inductive filter L_L

limits current i_{SC2} to set load current $I_L = 0.8$ amperes; see Fig. 3. With resistor $R_{SC2}=0$, the waveshape of secondary voltage e_{S1} is shown in Fig. 8(C).

Rectifier $RECT_1$ can be replaced by a fixed resistor, but rectifier $RECT_1$ prevents current from flowing in control resistor R_{SC2} during the interval of time the transistor is turned on (e_{S1} positive at the dot), thus reducing the power dissipated in resistor R_{SC2}.

Magnetic-Amplifier Control

The saturable current transformer transistor amplifier can be controlled by a small magnetic amplifier, Fig. 9(B) with the magnetic amplifier receiving its a-c power from the saturable current transformer transistor oscillator. Fig. 9(B) is the same circuit as that of Fig. 9(A) except that resistor R_{SC2} is replaced by saturable reactor SR_1. Reactor SR_1, combined with rectifier $RECT_1$, provides a half-wave self-saturating type of magnetic amplifier.[3,4] The magnetic amplifier controls the secondary current i_{SC2}, and controls the average load current I_L in a manner similar to that of the method resistor R_{SC2} of Fig. 9(A) in controlling the load current I_L, except that the magnetic amplifier phase controls current i_{SC2}.

The magnetic amplifier SR_1 of the Fig. 9(B) operates from the a-c voltage e_{S1} in a manner similar to that of a conventional self-saturating magnetic amplifier.[3,4] Different from a conventional magnetic amplifier, reactor SR_1 short-circuits its supply voltage e_{S1} when SR_1 saturates. When reactor SR_1 saturates, secondary voltage e_{S1} drops to about 5% of the voltage of e_{S1} when reactor SR_1 is unsaturated (typically, $e_{S1}=15$ volts with SR_1 unsaturated, and $e_{S1}=1.0$ volt with SR_1 saturated). The waveshapes of secondary voltage e_{S1} are shown in Fig. 10. When reactor SR_1 remains unsaturated for the entire cycle, load current $I_L=5$ amperes and voltage e_{S1} as shown in Fig. 10(A), which is the same as shown in Fig. 8(A). When reactor SR_1 remains saturated during the entire cycle of voltage e_{S1}, the waveshape of e_{S1} is shown in Fig. 10(C), which is similar to Fig. 8(C), and load current $I_L=0.8$ amperes. When reactor SR_1 is unsaturated at the start of the negative half-cycle (e_{S1} is negative at the dot in Fig. 9(B)), voltage e_{S1} starts at the same voltage shown in Fig. 10(A), and later in the cycle, when reactor SR_1 saturates, voltage e_{S1} drops to the voltage shown in Fig. 10(C). Fig. 10(B) shows the waveshape of voltage e_{S1} to provide load current $I_L=3$ amperes.

Control current I_{SC3}, Fig. 9(B), controls the time reactor SR_1 and saturates during the portion of the cycle that voltage e_{S1} is negative at the dot. The circuit of Fig. 9(B) is a combination of an oscillator and an amplifier that supplies high-frequency power (such as 1 kc to 10 kc) for a small magnetic amplifier, and provides a power amplifier stage. Fig. 11 shows the circuit of Fig. 9(B) except that a second stage is added to the magnetic amplifier. The magnetic amplifier of Fig. 11 is a 2-stage half-wave magnetic amplifier (commonly using a Supermalloy core for the first stage and an Orthonol or Deltamax core for the second stage) that controls the saturable current transformer transistor amplifier. The transfer characteristic of the circuit of Fig. 11 is shown in Fig. 12; the speed of response is shown in Fig. 13; and Fig. 14 shows the front and rear views of the circuit.

Reactive Loads

Reactive loads of both inductive and capacitive loads can be controlled by the magnetic-transistor amplifier. These reactive loads consist of control loads such as fields of motors or generators, solenoids, motor armatures, capacitive filtered loads, etc.

Inductive loads require a complementary rectifier (a "coasting" rectifier) as shown in Fig. 15 (switch S_L at position a) similar to that required for magnetic amplifiers. Motor armature loads and capacitive loads (S_L at positions b and c respectively of Fig. 15) require filter inductor L_F to avoid high peak voltage from being impressed across the load. When the average load voltage is low (minimum output power of the magnetic-transistor control), voltage peaks equal to the power source voltage, Fig. 4(D), occur across the load because of the very low resistance (sometimes 0.03Ω) of the transistor when it is turned on. Inductor L_F filters the load current, removing high peaks of voltage or current from being applied to the capacitive load. The complimentary rectifier (Fig. 15) is required to accompany inductor L_F due to the reaction of inductor L_F on the magnetic-transistor amplifier. Inductor L_F is normally small and economical oscillator frequency is commonly a kilocycle or greater.

Conclusions

The principle of using a combination oscillator and amplifier provides reliable, compact, fast-responding and economical control amplifiers. Operating at frequencies between one and ten kc permits the use of compact economical components for controlling power. The power dissipated in the saturable current transformer and the transistor used in the magnetic-transistor amplifier is approximately one per cent of the load power capacity of the amplifier, providing a high efficient economical amplifier. The multistage amplifier using only one transistor provides advantages of compactness and fast response exceeding that of conventional magnetic amplifiers.

Nomenclature

B = flux density in, kilogausses
C_L = load capacitor, microfarads (may be load filter or capacitive portion of load)
e_b = base-emitter voltage of transistor, volts, instantaneous
e_p = primary voltage of saturable current transformer, volts, instantaneous
e_S = secondary voltage of saturable current transformer, volts, instantaneous
H = magnetizing force of saturable current transformer core, oersteds
I_{BIAS} = bias current, amperes, average
i_L = load current, amperes, instantaneous
I_L = load current, amperes, average
i_S = secondary current of the saturable current transformer, amperes, instantaneous
i_{SC} = control signal current, amperes, instantaneous
I_{SC} = control signal current, amperes, average
L_F = filter inductance for the load, henrys
L_L = load inductance, henrys
N_S = number of turns on secondary winding of saturable current transformer
P_{RI} = primary winding of saturable current transformer
R_1, R_2 = resistance of resistors used to obtain bias for transistor, ohms
R_L = load resistance, ohms
R_{SC} = control signal circuit resistance, ohms
$RECT$ = rectifier
S_1 = switch to simulate chopper action of magnetic-transistor oscillator
S_L = load selector switch
SEC = secondary winding of saturable current transformer

References

1. A SINGLE-TRANSISTOR MAGNETIC-COUPLED OSCILLATOR, **Kan Chen, A. J. Schiewe.** *AIEE Transaction*, vol. 75, pt. I, Sept. 1956, pp. 396–400.

2. A VARIABLE FREQUENCY MAGNETIC-COUPLED MULTIVIBRATOR, **R. L. Van Allen.** *Ibid.*, vol. 74, pt. I, July 1955, pp. 356–61.

3. MAGNETIC-AMPLIFIER CIRCUITS (book), **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

4. MAGNETIC AMPLIFIERS (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955.

Discussion

T. Bernstein (University of Wisconsin, Madison, Wis.): The use of switching transistors in conjunction with saturating cores is receiving a good deal of attention at present. The paper is certainly a valuable and interesting addition. One of the points that is open to further discussion is the relatively high minimum output which occurs. The ratio of maximum to minimum output is approximately ten to one. Is this more or less a fundamental limit or can greater ranges be obtained? Has the use of a more square loop core material been investigated for this circuit?

R. E. Morgan: Thank you Dr. Bernstein for your excellent comments and questions. The high minimum output of this magnetic-transistor amplifier is disturbing for some applications. The minimum output can be decreased by design, but a sacrifice is made in speed of response.

Core materials of Orthonol, Deltamax, Supermalloy, *4-79* mopermalloy, mumetal, and silicon iron have been investigated. For simplicity, only one core material (*4-79* mopermalloy) was used in this paper. The choice of core material depends upon the application. Frequently, "chopping" frequencies above 5,000 cycles per second are used and *4-79* mopermalloy is a good general-purpose material for these frequencies. At lower frequencies (less than 1,000 cycles per second) silicon iron is good. The square-loop materials such as Orthonol and Deltamax present problems of "snap action" with these bistable **B-H** loops.

A Comprehensive Study of Magnetic-Amplifier Design

I—A Practical Design Method

R. M. HUBBARD
ASSOCIATE MEMBER AIEE

M. M. BISHOP
NONMEMBER AIEE

IN THE past, magnetic-amplifier design papers have been conspicuously absent from the literature. It would appear that most practical design methods are primarily empirical in nature and tend to be dependent on a notebook of past successful designs. At the other extreme are a few theoretical design methods involving such things as analytical approximations of the nonlinearities of the core materials. It is the purpose of this paper to introduce a new, practical design method which is the outgrowth of a study of magnetic amplifier theory using block-diagram techniques[1] and which, by its nature, involves an absolute minimum of "cut and try."

The application of this design method permits both the design of a minimum size amplifier to meet a given set of design requirements and the determination of the capabilities of an amplifier when the requirements are not completely specified. This paper presents the basic essentials of the design method which are independent of such things as the type of core material used, the signal waveform, etc.

The introduction of these latter complicating factors represents a refinement on the general design method. This refinement is normally taken care of by modifying the original specified requirements of a given amplifier to take into account, for example, the limited linear operating range characteristic of the selected core material and the change in gain caused by different signal waveforms. By circumventing these second-order complicating factors, the design can be generalized such that only two, limiting, static operating points (saturation and cutoff) and the response time need be considered.

Nomenclature

A_i = Intrinsic amplification factor $[(\bar{R})(N_c/N_x)^2/R_c]$
A_N = Window area occupied by a given winding
A_{Nc} = Window area occupied by control winding

Paper 58-792, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 8, 1958.

R. M. HUBBARD and M. M. BISHOP are with the Boeing Airplane Company, Seattle, Wash.

A_{Nx} = Window area occupied by load winding
A_w = Total usable window area of a core
a = Copper cross-sectional area of smallest usable wire
$\bar{B}$ = Average flux density
$\bar{B}_o$ = Average flux density at knee of a-c magnetization curve
D = Maximum allowable current density in wire
$\bar{E}$ = Average rectified voltage
$\bar{E}_c$ = Average rectified control (signal) voltage
E_d = Battery equivalent of diode threshold voltage
$\bar{E}_L$ = Average rectified load voltage
$(\bar{E}_L)_{co}$ = Average rectified load voltage at cut-off
$\bar{E}_{xo}$ = Average rectified voltage across N_x winding at knee of a-c magnetization curve
F = Fullness ratio of a given winding (ratio of A_N to A_w)
G_p = Amplifier power gain (ratio of $(\bar{V}\bar{A})_L$ to $(\bar{V}\bar{A})_{in}$)
$(G_p\eta_c)$ = Over-all power gain (ratio of $(\bar{V}\bar{A})_L$ to $(\bar{V}\bar{A})_c$)
$\bar{H}$ = Average magnetizing force
$\bar{H}_o$ = Average magnetizing force at knee of AC magnetization curve
HD = Minimum hole diameter in a fully wound toroidal core
$\bar{I}_c$ = Average rectified control current
$\bar{I}_L$ = Average rectified load current
$\bar{I}_R$ = Average rectified reset current
$\bar{I}_{xo}$ = Average rectified excitation current in N_x winding at knee of a-c magnetization curve
ID = Inside diameter of toroidal-core case
K = Internal voltage gain
K' = External voltage gain (ratio of $\Delta\bar{E}_L$ to $\Delta\bar{E}_c$)
k_R = Form factor of reset winding current
k_x = Form factor of load winding current
$\mathfrak{L}$ = Mean length magnetic path
M = Figure of merit $(G_p\eta_c/\tau_a - T)$
N = Number of turns in a given winding
N_c = Number of turns in a control winding
N_R = Number of turns in a reset winding
N_x = Number of turns in a load winding
OD = Outside diameter of toroidal-core case
Q_c = Ratio of reflected exciting impedance $\bar{R}(N_c/N_x)^2$ to control-circuit winding resistance (r_c)
Q_c^* = Ratio of reflected exciting impedance $\bar{R}(N_c/N_x)^2$ to resistance of the N_c winding (r_c^*)
Q_R^* = Minimum Q required for full control of the core by a reset winding operating at the rated current of the wire
Q_T^* = Maximum attainable Q corresponding to a winding which completely fills the usable window area A_w
Q_x = Ratio of exciting impedance $\bar{R}$ to load-circuit winding resistance r_x
Q_x^* = Ratio of exciting impedance $\bar{R}$ to resistance of the N_x winding r_x^*
$\bar{R}$ = Average exciting impedance of a core referred to the load (N_x) winding
$\bar{R}_N$ = Average exciting impedance referred to a given (N) winding
R_c = Total series resistance in the control circuit
r_c = Total series winding resistance in control circuit
r_c^* = Winding resistance of the control (N_c) winding
R_f = Forcing resistance in the control circuit
R_g = Signal generator internal resistance

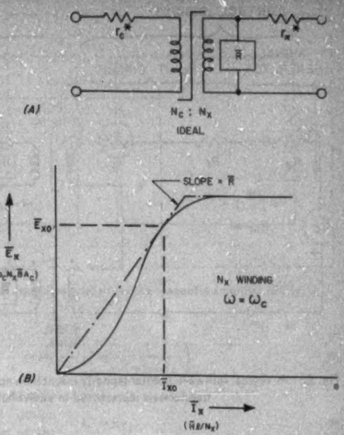


Fig. 1 (A). Saturable-transformer equivalent circuit (B). A-c magnetization curve for a typical saturable transformer

R_{IN} = Amplifier-input resistance as seen by the signal generator $(R_f + r_c)$
R_L = Effective load resistance (R_L' + diode dynamic resistance)
R_L' = True load resistance
R_s = Total signal-source resistance as seen by the amplifier $(R_g + R_f)$
r_w = Wire resistance per unit length
r_x = Total series winding resistance in load circuit
r_x^* = Winding resistance of the load (N_x) winding
S_1 = Length of shortest possible turn on a given core
S_2 = Length of longest possible turn on a given core
$\bar{S}_N$ = Mean length of turn of a given winding
$\bar{S}_T$ = Mean length of turn of a winding which completely fills usable window (A_w)
T = Time for one half-cycle of power supply frequency (ω_c)
$(\bar{V}\bar{A})_c$ = Average volt-amperes supplied by the signal generator
$(\bar{V}\bar{A}_{IN}$ = Average volt-amperes delivered to magnetic amplifier proper
$(\bar{V}\bar{A})_L$ = Average volt-amperes delivered to load
$(\bar{V}\bar{A})_x$ = Average core-excitation volt-amperes
α = Reset factor (Q_c^*/Q_R^*)
γ = Dynamic-range ratio $\Delta\bar{E}_L/(\bar{E}_L)_{co}$
η_c = Control circuit efficiency $(1/1 + R_s/r_c)$
η_L = Load circuit efficiency $(1/1 + r_x/R_L)$
ρ = Resistivity of wire
τ_a = Time for 63% response to a step change of control voltage
ω_c = Supply voltage angular frequency

The Saturable Transformer as the Basic Component of the Magnetic Amplifier

The basic element of which all magnetic amplifiers are built is the saturable transformer, which can be represented, see Fig. 1(A), in the equivalent circuit form previously introduced in the block-diagram analysis of magnetic amplifiers.[1] At the frequency of interest (ω_c), the nonlinear excitation impedance of the transformer is represented by an average resistance $(\bar{R})$ and the "ideal" transformer is taken as an ideal saturating-transformer. The resistance of the load (N_x) winding is r_x^* and that of the control (N_c) winding is r_c^*. Leakage reactance and winding capacitance are assumed negligible. The asterisks (*) are herein used to identify those parameters associated with the saturable transformer alone as against those parameters associated with the magnetic amplifier as a whole.

For purposes of this design method, the excitation impedance $(\bar{R})$ is assumed to have a single value which can be defined in terms of the normal a-c magnetization curve using the N_x winding. In Fig. 1(B) the average rectified voltage across the N_x winding, $\bar{E}_x$, is plotted versus the average rectified current that flows, $\bar{I}_x$. The only point of interest is that for "normal" excitation corresponding to the co-ordinates $(\bar{E}_{xo}, \bar{I}_{xo})$. Therefore, at the "knee" of the magnetization curve, the excitation impedance is

$$\bar{R} = \bar{E}_{xo}/\bar{I}_{xo} \tag{1}$$

and the core excitation in volt-amperes is

$$(\overline{VA})_x = (\bar{E}_{xo})(\bar{I}_{xo}) \tag{2}$$

The Q^* of a given winding is defined as the ratio of impedance to winding resistance, thus,

$$Q_x^* = \bar{R}/r_x^* \tag{3}$$

and

$$Q_c^* = \bar{R}(N_c/N_x)^2/r_c^* \tag{4}$$

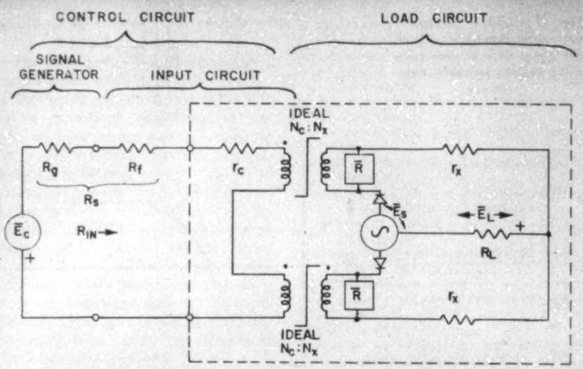


Fig. 2. A typical full-wave (center-tapped) magnetic amplifier with the component saturable transformers represented in equivalent circuit form

or

$$Q_c^* = \bar{R}'/r_c^* \tag{5}$$

The foregoing parameters associated with a particular transformer must be generalized for design purposes. Thus,

$$(\bar{R}_N/N^2) = (\bar{R}/N_x^2) = (\bar{R}'/N_c^2) = \omega_c\left[\frac{\bar{B}_o}{\bar{H}_o}\right]\left[\frac{A_c}{\mathcal{L}}\right] \tag{6}$$

In the foregoing, $\bar{B}_o$ and $\bar{H}_o$ are the average values of flux density and magnetizing force associated with operation under normally excited conditions (at $\omega = \omega_c$); whereas A_c is the net cross-sectional area of the core and $\mathcal{L}$ is the mean magnetic path length.

Using equation 6 and the following expression for the winding resistance of a given winding,

$$r_N = \frac{(\rho \bar{S}_N)}{(0.4A_N)}(N^2) \tag{7}$$

the Q_N^* of any winding is given by

$$Q_N^* = \frac{(0.4A_N)}{(\rho \bar{S}_N)}\left(\frac{1}{N}\right)^2\left\{(\omega_c)\left[\frac{\bar{B}_o}{\bar{H}_o}\right]\left[\frac{A_c}{\mathcal{L}}\right](N)^2\right\}$$

or

$$Q_N^* = \frac{(0.4A_N)}{(\rho \bar{S}_N)}(\omega_c)(\bar{B}_o/\bar{H}_o)(A_c/\mathcal{L}) \tag{8}$$

To simplify the succeeding work, sector winding on a toroidal core is assumed and, therefore, the mean length of turn ($\bar{S}_N$) of each winding is identical. The Q^* of a winding which completely fills the usable window area (A_w) is the maximum attainable and is designated Q_T^*. Thus,

$$Q_T^* = \frac{(0.4A_w)}{(\rho \bar{S}_T)}(\omega_c)\left[\frac{\bar{B}_o}{\bar{H}_o}\right]\left[\frac{A_c}{\mathcal{L}}\right] \tag{9}$$

The mean length of turn $\bar{S}_T$ for the fully wound core is characteristic of the dimensions of the particular core and has been formulated (see Appendix I) to permit the tabulation of Q_T^* for all standard toroidal cores.

A reset winding which is capable of fully resetting the core must have a Q^* which is equal to or greater than the minimum Q^* which corresponds to operation at the rated current of the wire in the reset winding. This minimum Q_R^* is given by (see Appendix I)

$$Q_R^* = (Q_T^*)\frac{(\bar{H}_o\mathcal{L})}{(0.4A_w)(D/k_R)} \tag{10}$$

In equation 10, k_R is the form-factor for the reset current, and D is the allowable current density in the wire. The ratio of the actual Q^* of a control winding, for example, to the minimum usable Q^* is designated the reset factor α. Thus,

$$\alpha = Q_c^*/Q_R^* \tag{11}$$

Since, for a given core, Q_N^* is proportional to the window area occupied by a given winding, then α is also the ratio of control-winding window area to minimum reset-winding window area. In addition, the ratio of α to the key parameter (Q_T^*/Q_R^*) is

$$\frac{\alpha}{Q_T^*/Q_R^*} = \frac{Q_c^*/Q_R^*}{Q_T^*/Q_R^*} = \frac{Q_c^*}{Q_T^*} \tag{12}$$

Clearly equation 12 is also the ratio of control-winding window area (A_{Nc}) to total usable window area (A_w). That is,

$$\frac{\alpha}{Q_T^*/Q_R^*} = \frac{A_{Nc}}{A_w} \tag{13}$$

The core excitation volt-amperes $(\overline{VA})_x$ may be generalized by noting that

$$(\overline{VA})_x = (\omega_c)(\bar{B}_o\bar{H}_o)(A_c\mathcal{L}) \tag{14}$$

and the per-turn voltage that a given winding can support under normally excited conditions is given by

$$\bar{E}/N = \omega\bar{B}_oA_c \tag{15}$$

The Design Parameters of a Typical Magnetic Amplifier

For purposes of illustrating this design method, the simple, full-wave, center-tapped magnetic amplifier of Fig. 2 has been selected as typical of the most common and most difficult to design class of magnetic amplifiers. The design problem resolves itself to the determination, from the magnetic amplifier, of the required values for the previously itemized saturable-transformer parameters and, from this, the selection of the optimum core.

In Fig. 2, the magnetic amplifier proper is defined as that circuitry enclosed by the dotted lines. Thus, the signal-generator resistance R_g and the forcing resistance R_f are external to the magnetic amplifier and their sum is considered to be the net signal source resistance R_s. The total winding resistance in the control circuit is designated r_c and the total winding resistance in the load circuit is r_x. Clearly the resistance of one saturable-transformer control winding r_c^* is one half of this total r_c, while the resistance of one saturable-transformer load winding r_x^* is equal to the total r_x. Note that as far as the signal

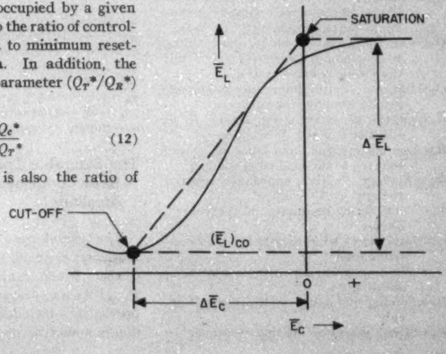


Fig. 3. A typical static transfer-characteristic for the circuit of Fig. 2

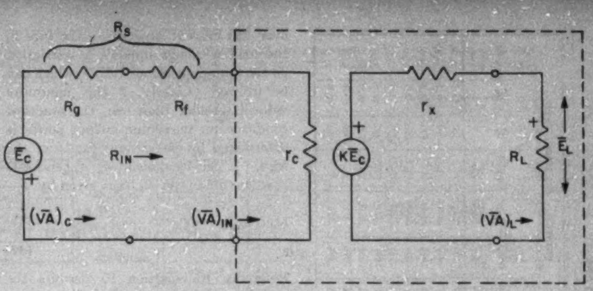


Fig. 4. Small-signal steady-state equivalent circuit for the magnetic amplifier of Fig. 2

generator is concerned, the input impedance of the magnetic amplifier (R_{in}) is the sum of the forcing resistance R_f and the winding resistance r_c.

A typical static transfer characteristic for the foregoing circuit is presented in Fig. 3. In line with the basic philosophy of this design technique, the requirements of a particular amplifier must be put in terms of the idealized transfer characteristic indicated by the dotted lines of Fig. 3. In other words, the requirements of the particular problem must be put into a form which is acceptable to the generalized design method. The transition from the typical, static transfer characteristic to the idealized one necessitates a familiarity with the characteristics of the core material to be utilized. Clearly, an extremely high order of accuracy is unobtainable in view of the relatively wide variation between the magnetic characteristics of different samples of the same core material. In fact, in the opinion of the authors, an attempt to design an amplifier such that the gain is solely determined by the characteristics of the core represents poor design techniques. That is, a certain amount of gain adjustment (e.g. negative feedback) is desirable to compensate for the normal variations in core characteristics.

Typically, the desired gain must be modified to account for the difference between the idealized transfer characteristic and the typical transfer characteristic at the same time that the ratio of gain to response time is held constant. In addition, the required dynamic range is modified to take into account the limited linear operating range which is characteristic of the particular core material.

In the following, the initial value of the control voltage is taken as zero such that $\Delta\bar{E}_c$ equals $\bar{E}_c$.

An important parameter associated with the static transfer characteristic is the dynamic-range ratio γ. This is defined by

$$\gamma=\Delta\bar{E}_L/(\bar{E}_L)_\infty \tag{16}$$

Of major importance in this design method is the small-signal steady-state equivalent circuit of Fig. 4. The key parameters, internal voltage gain K and 63% response time τ_a, are fixed by the circuit parameters according to the following expressions which were obtained from the previously mentioned block-diagram analysis.[1]

$$K=(A_i)(N_x/N_c) \tag{17}$$

$$\tau_a=(A_i+1)(\pi/\omega_c) \tag{18}$$

The intrinsic amplification factor A_i is defined as

$$A_i=\frac{\bar{R}(N_c/N_x)^2}{(R_s+r_c)} \tag{19}$$

Although the parameter A_i is valuable for purposes of analysis, this design method dictates the introduction of the parameter Q. If the Q of the control circuit is defined as

$$Q_c=\frac{\bar{R}'}{r_c}=\frac{\bar{R}(N_c/N_x)^2}{r_c} \tag{20}$$

and this is substituted into equation 19, the result is

$$A_i=(Q_c)\left(\frac{1}{1+R_s/r_c}\right) \tag{21}$$

If further, the efficiency of the control circuit η_c is defined as the ratio of the power delivered into the magnetic amplifier proper $(\overline{VA})_{in}$ to the total power drawn from the signal generator $(\overline{VA})_c$, it can be seen that

$$\eta_c=\frac{1}{1+R_s/r_c} \tag{22}$$

Thus, the intrinsic amplification factor is

$$A_i=Q_c\eta_c \tag{23}$$

the internal voltage gain is

$$K=(Q_c\eta_c)(N_x/N_c) \tag{24}$$

and the response time is

$$\tau_a=[(Q_c\eta_c)+1]\{\pi/\omega_c\} \tag{25}$$

The external voltage gain K' can be seen to be

$$K'=\frac{\Delta\bar{E}_L}{\Delta\bar{E}_c}=(K)\left(\frac{R_L}{R_L+r_x}\right)$$

or

$$K'=(K)\left(\frac{1}{1+r_x/R_L}\right) \tag{26}$$

If the efficiency of the load circuit η_L is defined as the ratio of load power to total power delivered by the amplifier, then

$$\eta_L=\left(\frac{1}{1+r_x/R_L}\right) \tag{27}$$

and

$$K'=K\eta_L \tag{28}$$

The various powers (average volt-amperes) can be determined by inspection to be as follows:

Load Power:

$$(\overline{VA})_L=\frac{(\Delta\bar{E}_L)^2}{R_L} \tag{29}$$

Signal Power:

$$(\overline{VA})_c=\frac{(\Delta\bar{E}_c)^2}{R_s+r_c} \tag{30}$$

Input Power:

$$(\overline{VA})_{in}=\frac{(\Delta\bar{E}_c\eta_c)^2}{r_c} \tag{31}$$

The foregoing may be normalized (see Appendix I) with respect to the core excitation volt-amperes to yield:

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=Q_x\left[\frac{(R_L/r_x)}{(1+R_L/r_x)^2}\right] \tag{32}$$

$$\frac{(\overline{VA})_c}{(\overline{VA})_x}=\frac{1}{(Q_c\eta_c)} \tag{33}$$

$$\frac{(\overline{VA})_{in}}{(\overline{VA})_x}=\frac{1}{Q_c} \tag{34}$$

It can readily be shown (see Appendix I) that the dynamic range ratio γ is given by

$$\gamma=Q_x/(1+R_L/r_x) \tag{35}$$

By inspection, the maximum voltage the transformer load winding (N_x) must support (at cutoff) is given by

$$\bar{E}_{x0}=\bar{E}_L/\eta_L \tag{36}$$

For maximum output power from a given amplifier at a given frequency, the load resistance (R_L) should be matched to the internal impedance (r_x). How-

TABLE I DESIGN PARAMETERS FOR TOROIDAL CORES (REFERENCE FREQUENCY 1000 CPS)

CORE NO	CASE DIMENSIONS			CORE DIMENSIONS		WINDING INFORMATION			HYMU-80, 4-79 MO-PERMALLOY, ETC. ($\bar{B}_0$ = 0.41 W/M²; $\bar{H}_0$ = 5.4 A-T/M)							ORTHONOL, HYPERNIK V, ETC. ($\bar{B}_0$ = 0.893 W/M²; $\bar{H}_0$ = 18.8 A-T/M)							MAGNESIL, SILECTRON, ETC. ($\bar{B}_0$ = 0.817 W/M²; $\bar{H}_0$ = 93 A-T/M)						
	OD IN	ID IN	H IN	$\mathcal{L}$ CM	A_c 1 MIL NET CM²	A_w IN²	S_1 IN	S_2 IN	E/N VOLTS/TURN	$\overline{(VA)}_x$ VOLT-AMP	R_N/N^2 OHM/TURN²	MIN$\left(\frac{R_L}{r_x}\right)$ OHM/OHM	Q_T^*	Q_R^*	MAX$\overline{(VA)}_L$ VOLT-AMP	E/N VOLTS/TURN	$\overline{(VA)}_x$ VOLT-AMP	$\bar{R}_N/N^2$ OHM/TURN²	MIN$\left(\frac{R_L}{r_x}\right)$ OHM/OHM	Q_T^*	Q_R^*	MAX$\overline{(VA)}_L$ VOLT-AMP	E/N VOLTS/TURN	$\overline{(VA)}_x$ VOLT-AMP	$\bar{R}_N/N^2$ OHM/TURN²	MIN$\left(\frac{R_L}{r_x}\right)$ OHM/OHM	Q_T^*	Q_R^*	MAX$\overline{(VA)}_L$ VOLT-AMP
12	.780	.382	.268	4.64	.1340	.055	.934	1.138	.034	.0086	.139	9.7	4340	21.4	3.32	.075	.065	.087	22.3	2720	46.6	7.42	.069	.294	.0161	20.3	500	42.7	6.76
13	.840	.412	.268	4.99	.1480	.072	.964	1.210	.038	.0103	.141	10.2	5520	22.5	4.80	.083	.078	.088	23.6	3450	49.1	10.7	.076	.352	.0164	21.5	640	44.9	9.80
14	.906	.443	.268	5.37	.1650	.092	.999	1.293	.042	.0123	.146	10.9	6930	23.8	6.86	.092	.093	.092	24.9	4340	51.9	15.3	.085	.423	.0169	22.8	800	47.5	13.9
15	.982	.475	.268	5.77	.1820	.113	1.043	1.349	.047	.0146	.150	11.6	8370	25.2	9.32	.102	.110	.094	26.5	5240	54.9	20.8	.093	.501	.0174	24.1	970	50.2	18.9
16	.954	.492	.325	5.75	.2160	.123	1.112	1.510	.055	.0173	.179	12.6	9900	27.3	12.1	.121	.131	.112	28.7	6200	59.4	26.8	.111	.592	.0207	26.2	1150	54.4	24.6
17	1.030	.530	.325	6.18	.2370	.155	1.150	1.614	.061	.0204	.183	13.2	12,100	28.4	16.7	.133	.154	.114	29.9	7560	61.8	37.2	.122	.699	.0212	27.3	1400	56.6	33.8
18	1.105	.580	.325	6.65	.2640	.196	1.175	1.735	.068	.0244	.189	14.0	15,000	30.0	23.6	.148	.185	.118	31.7	9400	65.4	52.4	.135	.838	.0220	29.0	1740	59.9	47.8
19	1.200	.610	.325	7.15	.292	.223	1.240	1.732	.075	.0290	.195	15.2	17,200	32.5	29.8	.164	.220	.122	34.4	10,770	70.8	66.0	.150	.996	.0225	31.4	2000	64.8	60.2
20	1.290	.655	.325	7.70	.322	.206	1.285	1.915	.083	.0345	.199	15.6	19,500	33.3	39.2	.180	.261	.125	35.3	12,200	72.6	86.8	.165	1.183	.0231	32.2	2260	66.4	79.4

ever, the current capacity of the wire in the load windings imposes a restriction on the minimum value of R_L/r_x which can be utilized. Clearly, if this minimum value is greater than one, the practical condition for maximum output power is determined by the current rating of the wire. It can be shown (see Appendix I) that the minimum R_L/r_x is given by

$$\text{Min}\,(R_L/r_x)=(Q_T^*)\frac{(\bar{H}_0\mathcal{L})}{(0.4A_w)(D/k_x)(2)}-1 \qquad (37)$$

Reference to equation 10 permits the simplification

$$\text{Min}'(R_L/r_x)=\left(\frac{Q_R^*}{2}\,\frac{k_x}{k_R}\right)-1$$

or, if k_R is assumed equal to k_x,

$$\text{Min}\,(R_L/r_x)=\frac{Q_R^*}{2}-1,\; k_R=k_x \qquad (38)$$

It is also informative to express such parameters as power gain, figure of merit, gain-bandwidth product, etc. in terms of Q. Therefore, the amplifier power gain is

$$G_p=\frac{\overline{(VA)}_L}{\overline{(VA)}_{in}}=Q_xQ_c\left[\frac{R_L/r_x}{(1+R_L/r_x)^2}\right] \qquad (39)$$

and the over-all power gain is

$$G_p\eta_c=\frac{\overline{(VA)}_L}{\overline{(VA)}_c}=Q_x(Q_c\eta_c)\left[\frac{R_L/r_x}{(1+R_L/r_x)^2}\right]$$

It is of particular interest to note that if the gain-bandwidth product and the figure of merit are defined as follows, simple expressions result.

$$\text{Gain bandwidth:}\quad K/(\tau_a-T)=(N_x/N_c)\left(\frac{1}{T}\right),\; T=\pi/\omega_c \qquad (40)$$

$$\text{Figure of merit:}\quad M=\frac{G_p\eta_c}{\tau_a-T}=\left[\frac{\overline{(VA)}_L}{\overline{(VA)}_x}\left(\frac{1}{T}\right)\right] \qquad (41)$$

At this point, the bulk of the necessary magnetic-amplifier design equations for a simple amplifier with no bias winding have been itemized. It will be seen that equations 11, 16, 25, 29, 30, 32, 33, and 35 contain sufficient information to permit the selection of the optimum core for a simple amplifier. In addition, equations 24, 26, 27, and 36 will be shown to be necessary for the selection of the number of turns on each winding.

Prior to the selection of the core for a particular amplifier, the foregoing information must be put in terms of saturable-transformer parameters rather than magnetic-amplifier parameters. This simply involves, for the subject circuit, the substitution

$$Q_c^*=2Q_c \qquad (42)$$

and

$$Q_x^*=Q_x \qquad (43)$$

Tabulation of Core Parameters

The heart of this design method is the tabulation of the important parameters for all of the standard core sizes and magnetic materials which are considered to be suitable for magnetic amplifier application. For simplicity, the tabulation lists the values of the various parameters at a reference frequency, 1,000 cycles per second. Appropriate correction factors permit the determination of the values at any other frequencies of interest.

Table I is an abstract of a more complete tabulation of the important core parameters and is based on a new proposal for standardizing core sizes, which is treated in the second part of this paper. All of the tabulated parameters have been previously treated herein with the exception of Max $\overline{(VA)}_L$. This parameter is the limiting value of output power which results when the available window is completely filled with load windings (i.e., $\alpha\rightarrow 0$) and the wire is operated at its rated current (i.e., $R_L/r_x=\text{Min}\,R_L/r_x$).

Thus,

$$\text{Max}\,\overline{(VA)}_L=\overline{(VA)}_x(Q_T^*)\times\left[\frac{\text{Min}\,(R_L/r_x)}{(1+\text{Min}\,(R_L/r_x))^2}\right] \qquad (44)$$

It should be pointed out that, although Table I is based on a new list of proposed standard cores, such a tabulation could be obtained for any toroidal cores. In fact, this design technique is applicable to any core (e.g. *D-U* laminations, *C* cores, etc.) which lend themselves to the calculations of the afore-mentioned design parameters.

The Mechanics of the Design Method

The actual mechanics of this design method can best be introduced by the use of the flow diagram of Fig. 5 and the corresponding block diagram of the necessary expressions, Fig. 6. The following steps in the design procedure can be visually traced by the use of these diagrams.

I. Given: Statement of the Problem

(A) Load circuit

1. The maximum swing in load voltage (ΔE_L) corresponding to operation from saturation to cutoff determines the required dynamic range.
2. The allowable load voltage at cutoff

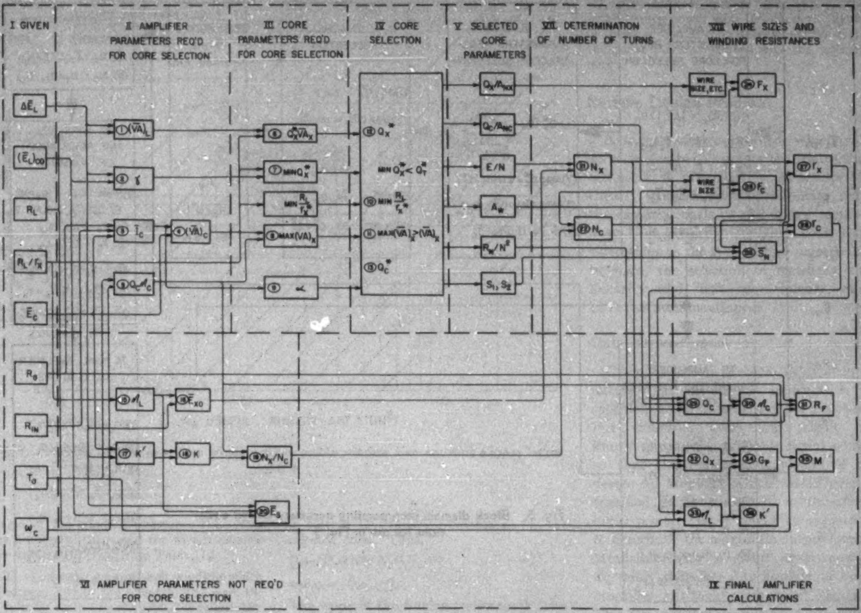

I GIVEN
II AMPLIFIER PARAMETERS REQ'D FOR CORE SELECTION
III CORE PARAMETERS REQ'D FOR CORE SELECTION
IV CORE SELECTION
V SELECTED CORE PARAMETERS
VII DETERMINATION OF NUMBER OF TURNS
VIII WIRE SIZES AND WINDING RESISTANCES
VI AMPLIFIER PARAMETERS NOT REQ'D FOR CORE SELECTION
IX FINAL AMPLIFIER CALCULATIONS
$\Delta\bar{E}_L$
$(\bar{E}_L)_{CO}$
R_L
R_L/r_X
$\bar{E}_C$
R_g
R_{IN}
T_a
ω_C
① $(\overline{VA})_L$
② γ
③ $\bar{I}_C$
④ $(\overline{VA})_C$
⑤ $Q_C\eta_C$
⑥ $Q_X^*\overline{VA}_X$
⑦ MIN Q_X^*
MIN $\frac{R_L}{r_X^*}$
⑧ MAX$(VA)_X$
⑨ α
⑫ Q_X^*
MIN $Q_X^* < Q_T^*$
⑩ MIN $\frac{R_L}{r_X^*}$
⑪ MAX$(\overline{VA})_X > (\overline{VA})_X$
⑬ Q_C^*
Q_X/A_{WX}
Q_C/A_{WC}
E/N
A_W
R_W/N^2
S_1, S_2
⑳ N_X
㉑ N_C
WIRE SIZE, ETC.
WIRE SIZE
F_X
F_C
$\bar{S}_N$
r_X
r_C
⑮ η_L
⑯ $\bar{E}_{XO}$
⑰ K'
⑱ K
⑲ N_X/N_C
㉙ E_S
Q_C
η_C
R_F
㉜ Q_X
㉞ G_P
㉟ M
η_L
K'

Fig. 5. Flow diagram for magnetic amplifier design

$(\bar{E}_L)_{\infty}$ affects the over-all efficiency but is normally determined by the nature of the desired static transfer characteristic.

3. The effective load resistance R_L includes the real load (R_L'), the dynamic forward resistance of the rectifier (see Appendix I), and the internal impedance of the carrier supply.

4. The ratio of effective load resistance to load-winding resistance (R_L/r_x) determines the regulation and the efficiency (η_L) of the load circuit.

(B) Control circuit

1. The full swing of control voltage $\Delta\bar{E}_c$ drives the amplifier from saturation to cutoff. (The addition of bias is a relatively simple matter which is herein omitted for the sake of brevity.)

2. R_g is the internal resistance of the signal generator.

3. The effective amplifier-input resistance (R_{in}) is the sum of the real steady-state input resistance r_c and the forcing resistance R_f. The input resistance should be stated as a minimum allowable value, since the required value of R_f may produce a higher R_{in} than specified.

4. The 63-per-cent response time (τ_a) has a minimum value which is limited by the time (T) for one-half cycle of the supply (carrier) frequency (ω_c).

5. The available carrier frequency is ω_c.

II. Amplifier Parameters Required for Core Selection

(A) The maximum signal power delivered to the effective load resistance is $(\overline{VA})_L$. The power dissipated at cutoff is normally wasted power.

(B) The required dynamic range ratio (γ) fixes a minimum allowable value on Q_x for a given design.

(C) The maximum control current $\bar{I}_c$ (or $\Delta\bar{I}_c$) is determined by the allowable input resistance (R_{in}), the signal-generator internal resistance (R_g) and the maximum control $\bar{E}_c$ (or $\Delta\bar{E}_c$). The total power supplied by the signal generator is merely the product of control voltage ($\bar{E}_c$) and control current ($\bar{I}_c$).

(D) The required response time (τ_a) and the carrier frequency (ω_c) fix the value of the important parameter $Q_c\eta_c$ (from equation 25).

III. The Core Parameters Required for Core Selection

The core parameters are simply determined from the preceding amplifier parameters.

(A) As previously mentioned:

$Q_x{}^* = Q_x$, $Q_c{}^* = 2Q_c$, $r_x{}^* = r_x$ and $r_c{}^* = r_c/2$.

(B) The product $Q_x{}^*(\overline{VA})_x$ is a constant which is fixed by the previously specified design requirements, namely, $(\overline{VA})_L$ and R_L/r_x (from equation 32).

(C) The minimum value of $Q_x{}^*$ was previously determined by the required γ.

(D) The maximum allowable value of the core excitation power $(\overline{VA})_x$ is determined by the available signal power $(\overline{VA})_c$ and the required response time (τ_a) (from equation 33).

(E) A reset winding which is capable of fully controlling the core must have a Q^* which is equal to or greater than the tabulated $Q_R{}^*$ corresponding to current limited operation of the winding. If $\bar{I}_c$ is equal to or greater than the rated current of the smallest wire which can be wound, then $Q_c{}^*$ is equal to $Q_R{}^*$ (i.e. $\alpha = 1$). However, if $\bar{I}_c$ is less than the rated current of the smallest wire which can be wound, then $Q_c{}^*$ must be greater than $Q_R{}^*$. In the latter case, the ratio of $Q_c{}^*$ to $Q_R{}^*$ (or α) is merely the ratio of the rated current of the smallest wire which can be used (i.e. aD/k_R) to the available control current ($\bar{I}_c$).

IV. Core Selection

(A) If the tabulated minimum $R_L/r_x{}^*$ is not equal to or less than the required $R_L/r_x{}^*$ which was specified in the original statement of the problem, then the tabulated value must be used and the resultant change in $Q_x{}^*(\overline{VA})_x$ calculated in section III(B).

(B) The tabulated value of $(\overline{VA})_x$ for the

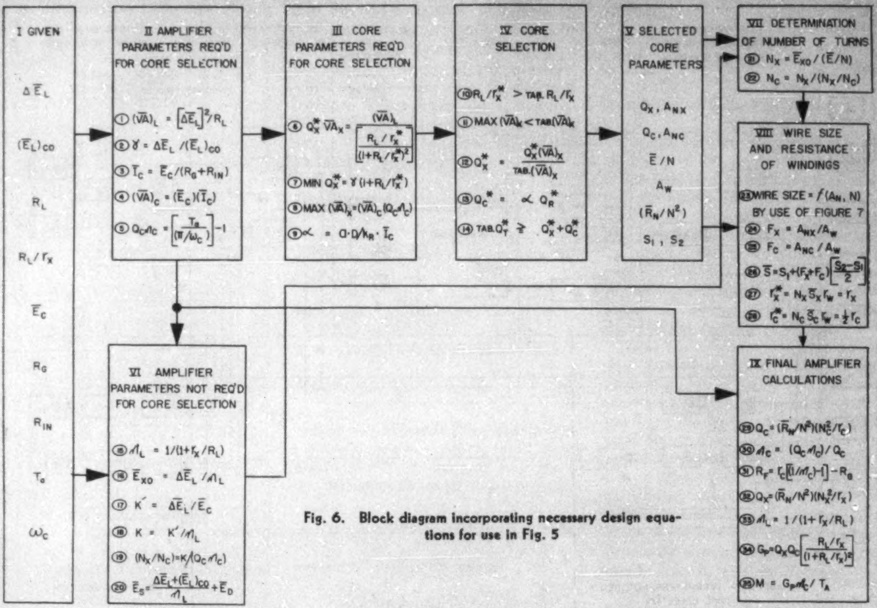

I GIVEN
$\Delta\bar{E}_L$
$(\bar{E}_L)_{co}$
R_L
R_L/r_x
$\bar{E}_c$
R_G
R_{IN}
T_a
ω_c
II AMPLIFIER PARAMETERS REQ'D FOR CORE SELECTION
① $(\overline{VA})_L = [\Delta\bar{E}_L]^2/R_L$
② $\gamma = \Delta\bar{E}_L/(\bar{E}_L)_{co}$
③ $\bar{I}_C = \bar{E}_C/(R_G+R_{IN})$
④ $(\overline{VA})_C = (\bar{E}_C)(\bar{I}_C)$
⑤ $Q_C\eta_C = \left[\frac{T_a}{(\pi/\omega_C)}\right]-1$
III CORE PARAMETERS REQ'D FOR CORE SELECTION
⑥ $Q_X^* \overline{VA}_X = \frac{(\overline{VA})_L}{\left[\frac{R_L/r_X^*}{(1+R_L/r_X^*)^2}\right]}$
⑦ MIN $Q_X^* = \gamma(1+R_L/r_X^*)$
⑧ MAX $(\overline{VA})_X = (\overline{VA})_C(Q_C\eta_C)$
⑨ $\alpha = a\cdot D/k_R\cdot\bar{I}_C$
IV CORE SELECTION
⑩ $R_L/r_X^* >$ TAB. R_L/r_X
⑪ MAX $(\overline{VA})_X <$ TAB$(\overline{VA})_X$
⑫ $Q_X^* = \frac{Q_X^*(\overline{VA})_X}{\text{TAB.}(\overline{VA})_X}$
⑬ $Q_C^* = \alpha Q_R^*$
⑭ TAB. $Q_T^* \geq Q_X^*+Q_C^*$
V SELECTED CORE PARAMETERS
Q_X, A_{NX}
Q_C, A_{NC}
$\bar{E}/N$
A_W
$(\bar{R}_N/N^2)$
S_1, S_2
VI AMPLIFIER PARAMETERS NOT REQ'D FOR CORE SELECTION
⑮ $\eta_L = 1/(1+r_X/R_L)$
⑯ $\bar{E}_{XO} = \Delta\bar{E}_L/\eta_L$
⑰ $K' = \Delta\bar{E}_L/\bar{E}_C$
⑱ $K = K'/\eta_L$
⑲ $(N_X/N_C) = K/(Q_C\eta_C)$
⑳ $\bar{E}_S = \frac{\Delta\bar{E}_L+(\bar{E}_L)_{co}}{\eta_L}+\bar{E}_D$
VII DETERMINATION OF NUMBER OF TURNS
㉑ $N_X = \bar{E}_{XO}/(\bar{E}/N)$
㉒ $N_C = N_X/(N_X/N_C)$
VIII WIRE SIZE AND RESISTANCE OF WINDINGS
㉓ WIRE SIZE = $f(A_N, N)$ BY USE OF FIGURE 7
㉔ $F_X = A_{NX}/A_W$
㉕ $F_C = A_{NC}/A_W$
㉖ $\bar{S} = S_1+(F_X+F_C)\left[\frac{S_2-S_1}{2}\right]$
㉗ $r_X^* = N_X\bar{S}_X r_W = r_X$
㉘ $r_C^* = N_C\bar{S}_C r_W = \frac{1}{2}r_C$
IX FINAL AMPLIFIER CALCULATIONS
㉙ $Q_C = (\bar{R}_N/N^2)(N_C^2/r_C)$
㉚ $\eta_C = (Q_C\eta_C)/Q_C$
㉛ $R_F = r_C[(1/\eta_C)-1]-R_G$
㉜ $Q_X = (\bar{R}_N/N^2)(N_X^2/r_X)$
㉝ $\eta_L = 1/(1+r_X/R_L)$
㉞ $G_P = Q_XQ_C\left[\frac{R_L/r_X}{(1+R_L/r_X)^2}\right]$
㉟ $M = G_P\eta_C/T_A$

Fig. 6. Block diagram incorporating necessary design equations for use in Fig. 5

selected core must be equal to or less than the maximum value determined in section III(D).

(C) Since $Q_x{}^*(\overline{VA})_x$ must be constant, $Q_x{}^*$ for a given core can be determined by dividing this constant by $(\overline{VA})_x$ for the given core. Clearly the resultant $Q_x{}^*$ must be less than the tabulated $Q_T{}^*$ by the amount $Q_c{}^*$ (where $Q_c{}^* = \alpha Q_R{}^*$). Good design dictates the selection of the smallest core which will meet the requirements outlined in both (B) and (C).

V. Selected-Core Parameters

(A) The core is capable of supporting $(\bar{E}/N)$-volts per turn under conditions of normal excitation.

(B) The usable window area (A_w) is that fractional part of the total window area which may be utilized in view of limitations of the winding machine.

(C) The area (A_N) occupied by a given winding can be determined by noting that winding area is a linear function of winding Q^*. Thus:

1. $A_{NX} = (Q_x{}^*/Q_T{}^*)A_w$
2. $A_{NC} = (Q_c{}^*/Q_T{}^*)A_w$

(D) The excitation impedance per turn squared $(\bar{R}_N/N^2)$ is a constant for the particular core.

(E) Utilization of the parameters S_1 and S_2 permits the calculation of the mean length of turn $(\bar{S}_N)$ of the windings.

1. The length of the shortest possible turn on the core (S_1) corresponds to the turn which is wound directly on the case.
2. The length of the longest possible turn on the core (S_2) corresponds to the turn which completely fills the usable window area A_w.

VI. Amplifier Parameters Not Required for Core Selection

(A) The load circuit efficiency (η_L) is fixed by the required R_L/r_x or by Min (R_L/r_x) of the selected core depending on which is larger.

(B) The maximum voltage $(\bar{E}_{xo})$ which must be supported by the load (N_x) winding corresponds to the maximum swing in load voltage. Thus, $\bar{E}_{xo} = \Delta\bar{E}_L/\eta_L$.

(C) The internal voltage gain K is merely the external voltage gain K' (i.e. $\Delta\bar{E}_L/\Delta\bar{E}_c$) divided by the load circuit efficiency η_L.

(D) The required-turns ratio (N_x/N_c) is fixed by the desired voltage gain K and by $Q_c\eta_c$ (via the desired τ_a).

(E) The supply voltage $\bar{E}_s$ is the sum of the following:

1. The maximum load voltage (at saturation), $\Delta\bar{E}_L + (\bar{E}_L)_{co}$, divided by the load circuit efficiency η_L.
2. The battery equivalent of the forward threshold voltage of the rectifier $\bar{E}_d$ (see Appendix I).

VII. Determination of Number of Turns

(A) Load winding

The number of turns (N_x) on the load winding is simply

$N_x = (\bar{E}_{xo})/(\bar{E}/N)$

(B) Control winding

The number of turns (N_c) on the control winding can be determined from the turns ratio (N_x/N_c) and N_x.

VIII. Determination of Wire Size and Winding Resistance

(A) Load winding

Using the number of load turns (N_x) and the allotted window area A_{NX}, the usable wire size may be determined from Fig. 7. The larger wire size of two choices should be used and any resultant change in A_{NX} noted for future use.

(B) Control winding

1. If α is greater than one, the control winding (N_c) should be wound with the previously determined smallest usable wire size. N_c turns of this size wire should fill the allotted window area (A_{NC}).
2. If α is equal to one, the number of control turns N_c and the allotted window area A_{NC} should be used to determine the wire size from Fig. 7. Any resultant change in A_{NC} should be noted for future use.

(C) Winding resistance

The determination of the resistance of each winding involves first the calculation of the mean length of turn S_N and then the calcu-

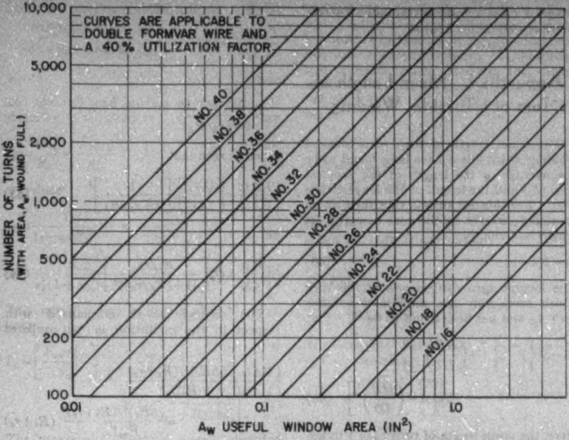


Fig. 7. Number of turns in a given winding versus usable window area assuming a space factor of 0.4

lation of the winding resistance (r_N) from the number of turns and the wire resistance per unit length (r_w) (from Table II).

1. Using the area of each winding which corresponds to the selected wire size, the fullness ratio ($F_N = A_N/A_w$) is calculated. Thus,

$F_x = A_{NX}/A_w$, $F_c = A_{NC}/A_w$

2. For sector winding, the mean length of turn of each winding is

$\bar{S}_N = S_1 + (F_x + F_c) \quad (S_2 - S_1)/2 = \bar{S}_x = \bar{S}_c$

3. The resistance of each winding ($r_N{}^*$) can be determined by substituting in the expression

$r_N{}^* = (N)(\bar{S}_N)(r_w)$

Thus:

$r_x{}^* = (N_x)(\bar{S}_x)(r_w)$, $r_c{}^* = (N_c)(\bar{S}_c)(r_w)$

Note that

$r_x = r_x{}^*$, $r_c = 2r_c{}^*$

IX. Final Calculations Using Actual Parameters of Selected Core

The actual values of several of the amplifier parameters can be compared with the design values by utilizing the foregoing information which applies to the selected core.

(A) The required forcing resistance can be calculated; or the experimental determination of this resistance is a simple matter, since at this point one is assured that the selected value will result in an amplifier which meets the initial design requirements.

The determination of R_f involves the calculation of the required value of control circuit efficiency (η_c) based on the actual Q_c and the value of ($Q_c\eta_c$), which was dictated by the required response time. Therefore, the actual value of Q_c (which will normally differ somewhat from the value determined in section III) must be calculated for the selected core. Thus,

$Q_c = (\bar{R}_N/N^2)(N_c{}^2)/r_c$

and $\eta_c = (Q_c\eta_c)/Q_c$

The value of the forcing resistance R_f can then be determined by using the definition of η_c (equation 22).

(B) The actual values of the load circuit efficiency (η_L) and the load circuit Q, (Q_x) can be determined by using the actual value of r_x which was determined in section VIII(C).

(C) The actual voltage gain (K'), the steady-state power gain (G_p) and the figure of merit M can also be determined by using the parameters associated with the selected core.

Sample Design Problem

A sample design problem which will serve to further illustrate the use of this technique is presented in Appendix II. Clearly, in the interest of simplicity, only the essential parts of the design method have been presented. Thus, such considerations as the type of core material to be used, the selection of rectifiers, the linearity and drift requirements, etc., have not been mentioned.

THE "OPEN-ENDED" DESIGN

Often, a design problem is of such a nature that insufficient information is specified in the statement of the problem. For example, everything might be known except the available signal power so that the requirement would be to design an amplifier which would give a specified performance with a minimum input power. Or it might be necessary to determine the maximum output power attainable with a given input power. Certainly there are any number of combinations of such open-ended designs. This design technique, in slightly modified form, lends itself to the optimization of such designs. For example, it has been seen that

$$\frac{(\overline{VA})_L}{(\overline{VA})_x} = Q_x\left[\frac{R_L/r_x}{(1+R_L/r_x)^2}\right] \tag{32}$$

Table II. Wire Table

B & S Gage	Diameter, Mils	Area, Circular Mils	Area, Square Inch	Ohms per 1,000 Feet at 25 Degrees C	RMS Current-Carrying Capacity,* Amperes
16	50.82	2,583	0.002028	4.016	5.17
17	45.26	2,048	0.001609	5.064	4.10
18	40.30	1,624	0.001276	6.385	3.25
19	35.89	1,288	0.001012	8.051	2.58
20	31.96	1,022	0.0008023	10.15	2.04
21	28.46	810.1	0.0006363	12.80	1.62
22	25.35	642.4	0.0005046	16.14	1.28
23	22.57	509.5	0.0004002	20.36	1.02
24	20.10	404.0	0.0003173	25.67	0.808
25	17.90	320.4	0.0002517	32.37	0.641
26	15.94	254.1	0.0001996	40.81	0.508
27	14.20	201.5	0.0001583	51.47	0.403
28	12.64	159.8	0.0001255	64.90	0.320
29	11.26	126.7	0.00009953	81.83	0.253
30	10.03	100.5	0.00007894	103.2	0.201
31	8.928	79.70	0.00006260	130.1	0.159
32	7.950	63.21	0.00004964	164.1	0.126
33	7.080	50.13	0.00003937	206.9	0.100
34	6.305	39.75	0.00003122	260.9	0.0795
35	5.615	31.52	0.00002476	329.0	0.0630
36	5.000	25.00	0.00001964	414.8	0.0500
37	4.453	19.83	0.00001557	532.1	0.0397
38	3.965	15.72	0.00001235	659.6	0.0314
39	3.531	12.47	0.000009793	831.8	0.0249
40	3.145	9.88	0.000007766	1,049	0.0198

* At 2 milliamperes per circular mil.

$$M=\left[\frac{(\overline{VA})_L}{(\overline{VA})_x}\right]\left[\frac{1}{T}\right] \tag{41}$$

and

$$G_p=Q_xQ_c\left[\frac{R_L/r_x}{(1+R_L/r_x)^2}\right] \tag{39}$$

If the core window area is to be fully utilized it can be seen that

$$Q_x{}^*=Q_T{}^*-Q_c{}^*=Q_x$$

and

$$Q_c{}^*=2Q_c.$$

Therefore

$$(\overline{VA})_L\propto(Q_T{}^*-Q_c{}^*),$$
$$M\propto(Q_T{}^*-Q_c{}^*)$$

and

$$G_p\propto(Q_c{}^*/2)(Q_T{}^*-Q_c{}^*)$$

Clearly, to maximize $(\overline{VA})_L$ and M, the window area should be completely filled by the load winding (i.e. $Q_c{}^*\rightarrow 0$).

However, to maximize G_p it can be seen that

$$\frac{\partial G_p}{\partial Q_c{}^*}=0=Q_T{}^*-2Q_c{}^*$$

$$\therefore Q_c{}^*=Q_T{}^*/2$$

Thus, the window should be half filled by the control winding to maximize the power gain G_p. The application of the other design equations permits similar general conclusions to be drawn as an aid to the solution of this particular type of design problem.

The list of proposed standard cores, of which Table I is a partial abstract, is particularly valuable when the design requirements are incompletely specified. This fact is clarified by Part II of this paper which describes the performance parameters which are inherent to the tabulation.

In effect, the core tabulation contains information as to the relative performance of the tabulated cores rather than merely the physical dimensions.

Conclusions

The basic principles of a new magnetic-amplifier design technique based on a study of magnetic amplifier theory have been presented. This technique involves an absolute minimum of purely empirical data but, at the same time, avoids impractical analytical studies of the nonlinearities inherent to the component magnetic elements. Although a typical full-wave magnetic amplifier circuit has been taken as a model for this presentation, the design technique can readily be extended to half-wave magnetic amplifiers, toroidal transformers, and a variety of other magnetic circuits.

Appendix I. Mean Length of Turn of a Toroidal Winding

The mean length of turn of a toroidal winding which fills A_w has been found to be the average of the length of the shortest turn (S_1) and the length of the longest turn (S_2). The shortest turn is that turn which is wound directly on the core case and its length is obviously,

$$S_1=2(h+b) \tag{45}$$

The longest turn corresponds to the turn which completely fills the usable window area A_w and has been found to be

$$S_2=S_1\left\{1-2\left[1-\left(\frac{HD}{ID}\right)^2\right]\right\}+2\bar{S}_0\left[1-\left(\frac{HD}{ID}\right)^2\right] \tag{46}$$

Here, the mean length of turn $\bar{S}_0$ of a hypothetical winding which completely fills the window area (no hole) is given by

$$\bar{S}_0=2h\left[1+\frac{(0.98)(ID/OD+0.306)}{(1-ID/OD)(h/b)}\right], \quad ID/OD\leq 0.9 \tag{47}$$

Formulation of Minimum Q^* of a Reset Winding

The reset current required for full control is given by

$$\bar{I}_R=\bar{H}_o\mathcal{L}/N_R \tag{48}$$

and the maximum allowable current in the winding is

$$\begin{aligned}\text{Max } \bar{I}_R&=(D/k_R)(a_{\text{wire}})\\&=(D/k_R)(0.4A_{NR}/N_R)\end{aligned} \tag{49}$$

Therefore, under current limited conditions,

$$\bar{H}_o\mathcal{L}/N_R=(D/k_R)(0.4A_{NR}/N_R) \tag{50}$$

and

$$A_{NR}=\frac{\bar{H}_o\mathcal{L}}{(D/k_R)(0.4)} \tag{51}$$

But, since

$$\frac{Q_N{}^*}{A_N}=\frac{Q_T{}^*}{A_w}=\frac{Q_R{}^*}{A_{NR}}$$

it can be seen that

$$Q_R{}^*=Q_T{}^*\left[\frac{\bar{H}_o\mathcal{L}}{(D/k_R)(0.4A_w)}\right] \tag{52}$$

Core Excitation Volt-Amperes as a Normalizing Factor

The output power normalized with respect to the excitation volt-amperes is as follows

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=\frac{(\Delta\bar{E}_L)^2/R_L}{\bar{E}_{xo}\bar{I}_{xo}}=\frac{(\bar{E}_{xo})^2(\eta_L)^2/R_L}{\bar{E}_{xo}\bar{I}_{xo}}=\frac{(\bar{E}_{xo}/\bar{I}_{xo})(\eta_L)^2}{R_L}$$

Since

$$\frac{\bar{E}_{xo}/\bar{I}_{xo}}{r_x}=Q_x$$

and

$$\eta_L=\frac{1}{1+(r_x/R_L)}$$

it can be seen that

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=\left[\frac{Q_xr_x}{R_L}\right]\left[\frac{1}{1+r_x/R_L}\right]^2 \tag{53}$$

or

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=Q_x\left[\frac{R_L/r_x}{[1+(R_L/r_x)]^2}\right] \tag{54}$$

The control power normalized with respect to the excitation in volt-amperes is

$$\frac{(\overline{VA})_c}{(\overline{VA})_x}=\frac{(\Delta E_c)^2/(R_s+r_c)}{\bar{E}_{xo}\bar{I}_{xo}}=\frac{(\Delta E_c)^2/(R_s+r_c)^2}{\bar{E}_{xo}\bar{I}_{xo}}\,(R_s+r_c)$$

But,

$$\frac{\Delta E_c}{R_s+r_c}=\frac{\bar{H}_o\mathcal{L}}{N_c}=(N_x/N_c)(\bar{I}_{xo})$$

and

$$\eta_c=\frac{1}{1+R_s/r_c}$$

Therefore,

$$\frac{(\overline{VA})_c}{(\overline{VA})_x}=\frac{[(N_x/N_c)(\bar{I}_{xo})]^2(r_c)}{(\bar{E}_{xo}\bar{I}_{xo})(\eta_c)} \tag{55}$$

or

$$\frac{(\overline{VA})_c}{(\overline{VA})_x}=\frac{1}{Q_c\eta_c} \tag{56}$$

The input power normalized with respect to the excitation in volt-amperes is

$$\frac{(\overline{VA})_{IN}}{(\overline{VA})_x}=\frac{[(\bar{I}_{xo})(N_x/N_c)]^2r_c}{\bar{E}_{xo}\bar{I}_{xo}} \tag{57}$$

or

$$\frac{(\overline{VA})_{IN}}{(\overline{VA})_x}=\frac{1}{Q_c} \tag{58}$$

Formulation of the Dynamic Range Ratio

Since

$$\gamma=\frac{\Delta\bar{E}_L}{(\bar{E})_{L\infty}}=\frac{\bar{E}_{xo}/\eta_L}{\bar{I}_{xo}R_L}$$

$$Q_x=(\bar{E}_{xo}/\bar{I}_{xo})/r_x$$

and

$$\eta_L=\frac{1}{1+r_x/R_L}$$

it can be seen that

$$\gamma=Q_x/(1+R_L/r_x) \tag{59}$$

FORMULATION OF THE MINIMUM VALUE OF R_L/r_x

For a practical design

$$\gamma >> 1$$

and the full load current at saturation is given by

$$I_L = \frac{\Delta\bar{E}_L}{R_L} = \frac{E_{x0}\eta_L}{R_L}$$

or; since

$$\eta_L = \frac{1}{1+r_x/R_L}$$

$$I_L = \frac{E_{x0}}{r_x(1+R_L/r_x)} \tag{60}$$

This may be put in terms of Q_x^* as follows:

$$I_L = \left[\frac{E_{x0}/I_{x0}}{r_x}\right]\left[\frac{I_{x0}}{1+R_L/r_x}\right] = Q_x^*\left[\frac{(\bar{H}_o\mathcal{L})/N_x}{1+R_L/r_x}\right] \tag{61}$$

The maximum allowable current dictated by the current capacity of the wire is as follows

$$\text{Max } I_L = 2(D/k_x)(a_{\text{wire}}) = 2(D/k_x)(0.4A_{Nx}/N_x) \tag{62}$$

Therefore,

$$Q_x\left[\frac{\bar{H}_o\mathcal{L}/N_x}{1+\text{Min}(R_L/r_x)}\right] = 2(D/k_x)(0.4A_{Nx}/N_x) \tag{63}$$

and

$$1+\text{Min }(R_L/r_x) = Q_x\left[\frac{\bar{H}_o\mathcal{L}}{(D/k_x)(0.4A_{Nx})^*}\right] \tag{64}$$

Since

$$\frac{Q_x}{A_{Nx}} = \frac{Q_x^*}{A_{Nx}} = \frac{Q_T^*}{A_w}$$

it can be seen that

$$\text{Min }(R_L/r_x) = Q_T^*\left[\frac{\bar{H}_o\mathcal{L}}{(D/k_x)(0.4A_w)(2)}\right] - 1 \tag{65}$$

RECTIFIER EQUIVALENT CIRCUIT

It has been found valuable to simplify the nonlinear characteristics of magnetic-amplifier rectifiers by assuming the equivalent circuit to be an equivalent battery (of approximately 0.5 to 0.7 volt) with an internal impedance corresponding to the dynamic forward resistance of the rectifier. This simplification is particularly valid in the case of silicon fused-junction diodes.

Appendix II. Sample Design Problem

The actual (practical) design requirements must be put into a form which is acceptable to this design method in terms of the idealized static transfer-characteristics. The utilization of such a modification of the practical design requirements is typified by the sample design problem which follows. Clearly, the sample problem cannot include all of the practical considerations which are based on knowledge and common sense, therefore, this design method is merely intended to serve as a valuable guide rather than as a "cook book."

I. Given

$\Delta\bar{E}_L = 23$ volts
$(\bar{E}_L)_{\infty} = 0.1$ volt
$R_L = 115$ ohms

$R_L/r_x = 30$
$\omega_c = 6{,}280$ radians per second
$\bar{E}_c = 2$ volts

$R_g = 250\ \Omega$
$R_{IN} = 250\ \Omega$ (min)
$\tau_a = 0.002$ second

II. Amplifier Parameters Required for Core Selection

$$(\overline{VA})_L = \frac{(\Delta\bar{E}_L)^2}{R_L} = \frac{(23)^2}{(115)} = 4.60 \text{ volt-amperes}$$

$$\gamma = \frac{\Delta\bar{E}_L}{(\bar{E}_L)_{\infty}} = \frac{(23)}{(0.1)} = 230$$

$$\text{Min } Q_x = \gamma(1+R_L/r_x) = (230)(1+30) = 7{,}130$$

$$I_c = \frac{\bar{E}_c}{R_g+R_{IN}} = \frac{(2)}{(250+250)}\ 0.004 \text{ ampere}$$

$$(\overline{VA})_c = \bar{E}_cI_c = (2)(0.004) = 0.008 \text{ volt-ampere}$$

$$Q_c\eta_c = \left[\frac{\tau_a}{\pi/\omega_c} - 1\right] = \left[\frac{0.002}{\pi/6280} - 1\right] = 3$$

III. Core Parameters Required for Core Selection

$$Q_x(\overline{VA})_x = \frac{(\overline{VA})_L}{\left[\dfrac{R_L/r_x}{(1+R_L/r_x)^2}\right]} = \frac{4.60}{\left[\dfrac{30}{(31)^2}\right]} = 147.5$$

$$\text{Max } (\overline{VA})_x = (\overline{VA})_c(Q_c\eta_c) = (0.008)(3) = 0.024 \text{ volt-ampere}$$

If *#40* is the smallest wire which can be wound,

$$\alpha = (a)(D)/(k_R)(I_c) = (7.76\times10^{-6})(2{,}546)/(1.11)(0.004) = 4.45$$

IV. Core Selection

A study of the core tabulation, Table I, reveals that *HYMU-80* core *#16* most closely fits the foregoing requirements. Thus:

The required $R_L/r_x(30)$ is greater than the tabulated minimum $R_L/r_x(12.7)$, and the maximum allowable $(\overline{VA})_x(0.024)$ is greater than the actual $(\overline{VA})_x(0.0173)$.

$$\text{Required } Q_x^* = \frac{Q_x(\overline{VA})_x}{\text{TAB}(\overline{VA})_x} = \frac{147.5}{0.0173} = 8{,}518$$

Since the minimum $Q_x^*(7{,}130)$ dictated by the required value of γ is less than the foregoing $Q_x^*(8{,}518)$, γ is not a limiting factor.

$$Q_c^* = \alpha Q_R^* = (4.45)(27.3) = 121$$

Since the required Q_T^* is the sum of Q_x^* and Q_c^*, $Q_T^* = (8{,}518)+(121) = 8{,}639$, and this value is less than the tabulated $Q_T^*(9{,}900)$.

In order to determine the required wire sizes, the area occupied by each winding must be determined as follows:

$$A_{Nx} = \frac{Q_x^*}{Q_T^*}A_w = \frac{(8{,}518)}{(9{,}900)}(0.123) = 0.106 \text{ sq. in.}$$

$$A_{Nc} = \frac{Q_c^*}{Q_T^*}A_w = \frac{(121)}{(9{,}900)}(0.123) = 0.0015 \text{ sq.in.}$$

VI. Amplifier Parameters Required for Core Selection

$$\eta_L = \frac{1}{\left(1+\dfrac{r_x}{R_L}\right)} = \frac{1}{\left(1+\dfrac{1}{30}\right)} = 0.968$$

$$\bar{E}_{x0} = \Delta\bar{E}_L/\eta_L = (23/0.968) = 23.76 \text{ volts}$$

$$K = \frac{\Delta\bar{E}_L/\eta_L}{\bar{E}_c} = \frac{(23/0.968)}{(2)} = 11.88 \text{ volts per volt}$$

$$(N_x/N_c) = \frac{K}{(Q_c\eta_c)} = \frac{(11.88)}{(3)} = 3.96 \text{ turns per turn}$$

$$\bar{E}_s = \frac{\Delta\bar{E}_L+(\bar{E}_L)_{\infty}}{\eta_L} + \bar{E}_d = \frac{(23+0.1)}{(0.968)} + (0.7) = 24.6 \text{ volts}$$

VII. Determination of Number of Turns

$$N_x = \bar{E}_{x0}/\left(\frac{E}{N}\right) = \frac{(23.76)}{(0.055)} = 432 \text{ turns}$$

$$N_c = N_x/\left(\frac{N_x}{N_c}\right) = \frac{(432)}{(3.96)} = 109 \text{ turns}$$

VIII. Wire Size and Resistance of Windings

Using the turns per area curves (Fig. 7) and areas from section IV,

$N_x = 432$ turns, if *#29* wire, $A_{Nx} = 0.11$ sq. in.

$N_c = 109$ turns, if *#40* wire, $A_{Nc} = 0.0022$ sq. in.

The difference between actual areas and those found under section IV is due to the finite number of wire sizes available.

$$F_x = \frac{A_{Nx}}{A_w} = \frac{0.11}{0.123} = 0.895$$

$$F_c = \frac{A_{Nc}}{A_w} = \frac{0.0022}{0.123} = 0.0179$$

$$\bar{S}_n = S_1 + (F_x+F_c)\left(\frac{S_2-S_1}{2}\right) = (1.112) + (0.913)\left(\frac{1.510-1.112}{2}\right) = 1.294 \text{ inches}$$

$$r_x^* = N_x\bar{S}r_w = (432)(1.294)\left(\frac{0.0818}{12}\right) = 3.79\ \Omega,\ (r_x = r_x^*)$$

$$r_c^* = N_c\bar{S}r_w = (109)(1.294)\left(\frac{1.049}{12}\right) = 12.3\ \Omega,\ (r_c = 2r_c^* = 24.6\ \Omega)$$

IX. Final Amplifier Calculations

Because of the practical limitations on wire sizes, the actual parameters become

$$Q_c=(\bar{R}_N/N^2)\frac{N_c^2}{r_c}=(0.179)(109)^2/(24.6)$$
$$=86.5$$

$$\eta_c=\frac{(Q_c\eta_c)}{Q_c}=\frac{(3)}{(86.5)}=0.0347$$

$$R_F=r_c\left(\frac{1}{\eta_c}-1\right)-R_g=$$
$$(24.6)\left(\frac{1}{0.0347}-1\right)-250=434\text{ ohms}$$

$$Q_z=(R_N/N^2)(N_z^2/r_z)$$
$$=(0.179)(432)^2/(3.79)=8{,}814$$

$$\eta_L=\frac{1}{(1+r_z/R_L)}=\frac{1}{\left(1+\frac{3.79}{115}\right)}$$
$$=0.968,\ \frac{R_L}{r_z}=\frac{115}{3.79}=30.34$$

$$G_p=Q_zQ_c\left[\frac{R_L/r_z}{(1+R_L/r_z)^2}\right]$$
$$=(8{,}814)(86.5)\left[\frac{30.34}{(31.34)^2}\right]=29{,}650$$

$$M=\frac{G_p\eta_c}{r_a}=\frac{(29{,}650)(0.0347)}{(0.002)}=516{,}500$$

It is interesting to note that the value of R_F which was previously determined yields a higher input resistance than the required value. Thus, although the amplifier performance meets the specifications, the actual required signal power is less than that specified.

Reference

1. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, Robert M. Hubbard. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

A Comprehensive Study of Magnetic-Amplifier Design

II—Toroidal-Core Dimensional Ratios as Factor in Magnetic-Amplifier Design

R. M. HUBBARD
ASSOCIATE MEMBER AIEE

M. M. BISHOP
NONMEMBER AIEE

THE APPLICATION of a block-diagram method of analysis for magnetic amplifiers[1] to practical design problems has produced a design technique which involves a minimum of cut and try while yielding near-optimum results. An indispensable part of this design method is a tabulation of the physical constants of the available toroidal magnetic cores. In the process of tabulating standard cores, it became clear to the authors that there was absolutely nothing systematic about the manner in which standard core sizes had evolved in the past. Clearly, with one exception,[2] little attempt had been made to determine the optimum dimensional ratios or to eliminate overlaps and gaps in the power ratings of the cores. It would appear that magnetic amplifiers were destined to suffer the same fate as vacuum-tube amplifiers as a result of this random development of core sizes. Therefore, an investigation to determine the effect of toroidal-core dimensional ratios on magnetic-amplifier performance was undertaken. Subsequent to the initiation of this investigation the first proposal for standardizing toroidal-core sizes was published by the AIEE Committee on Magnetic Amplifiers.[3]

This paper is intended to present a logical method by which the optimum toroidal-core configurations can be determined and also to propose a list of standard core sizes which effectively covers the practical range of power levels with a minimum number of cores. The proposed AIEE Standards on recommended tape widths and limiting values of dimensional ratios are adhered to wherever possible.

Nomenclature

A_c = net cross-sectional area of core
A_N = window area occupied by a given winding
A_{Nc} = window area occupied by control winding
A_{Nz} = window area occupied by load winding
A_w = total usable window area of core (herein, total window area)
$\bar{B}$ = average flux density
$\bar{B}_0$ = average flux density at knee of a-c magnetization curve
b = core build-up
D = maximum allowable current density in wire
$\bar{E}$ = average rectified voltage
$\bar{E}_c$ = average rectified control (signal) voltage
$\bar{E}_d$ = battery equivalent of diode threshold voltage
$\bar{E}_L$ = average rectified load voltage
$\bar{E}_{z0}$ = average rectified voltage across N_z winding at knee of a-c magnetization curve
G_p = amplifier power gain (ratio of $(\bar{V}\bar{A})_L$ to $(\bar{V}\bar{A})_{IN}$)
$(G_p\eta_c)$ = over-all power gain (ratio of $(\bar{V}\bar{A})_L$ to $(\bar{V}\bar{A})_c$)
$\bar{H}$ = average magnetizing force
$\bar{H}_0$ = average magnetizing force at knee of a-c magnetization curve
h = core height
h/b = ratio of core height to build-up
$\bar{I}_c$ = average rectified control current
$\bar{I}_L$ = average rectified load current
ID/OD = ratio of core inside diameter (ID) to outside diameter (OD)
k_R = form factor of reset current
k_z = form factor of load current
l = mean length of magnetic path
M = figure of merit $(G_p\eta_c)/(\tau_a-T)$
N = number of turns in a given winding
N_c = number of turns in control winding
N_z = number of turns in load winding
Q_b = ratio of reflected exciting impedance $\bar{R}(N_b/N_z)^2$ to bias circuit winding resistance r_b
$Q_b{}^*$ = ratio of reflected exciting impedance $\bar{R}(N_b/N_z)^2$ to resistance of the N_b winding $r_b{}^*$
Q_c = ratio of reflected exciting impedance $\bar{R}(N_c/N_z)^2$ to control circuit winding resistance r_c
$Q_c{}^*$ = ratio of reflected exciting impedance $\bar{R}(N_c/N_z)^2$ to resistance of the N_c winding $r_c{}^*$
$Q_n{}^*$ = the Q of a given winding
$Q_r{}^*$ = minimum Q required for full control of core by reset winding operating at rated current of the wire
$Q_T{}^*$ = the Q of winding which completely fills the usable window area (A_w)
Q_z = ratio of exciting impedance $\bar{R}$ to load circuit winding resistance r_z
$Q_z{}^*$ = ratio of exciting impedance $\bar{R}$ to resistance of the N_z winding $r_z{}^*$
$\bar{R}$ = average exciting impedance of core referred to the load (N_z) winding
R_c = total series resistance in control circuit
r_c = total series winding resistance in control circuit
$r_c{}^*$ = winding resistance of the control (N_c) winding
R_f = forcing resistance in control circuit
R_g = signal generator internal resistance
$r_n{}^*$ = winding resistance of a given winding
R_s = total signal source resistance as seen by the amplifier (R_g+R_f)
r_w = wire resistance per-unit length
r_z = total series winding resistance in load circuit

Paper 58-793, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 8, 1958.

R. M. HUBBARD and M. M. BISHOP are with the Boeing Airplane Company, Seattle, Wash.

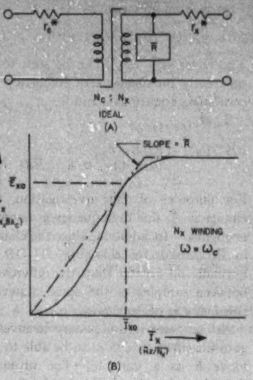


Fig. 1 (left). Saturable-transformer equivalent circuit (A). A-c magnetization curve for typical saturable transformer (B)

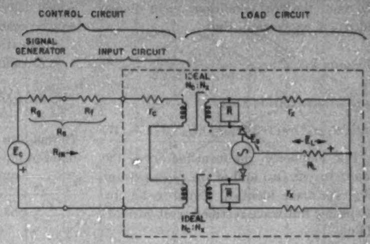


Fig. 2 (right). A typical full-wave (center-tapped) magnetic amplifier with component saturable transformers represented in equivalent-circuit form

r_x^* = winding resistance of the load (N_x) winding
S = length of one turn of a given winding
S_1 = length of shortest possible turn on a given core
S_2 = length of longest possible turn on a given core
$\bar{S}_n$ = mean length of turn of a given winding
$\bar{S}_0$ = mean length of turn of a winding which completely fills the core window
T = time for one half-cycle of power supply frequency (ω_c)
$(\overline{VA})_c$ = average volt-amperes supplied by signal generator
$(\overline{VA})_{IN}$ = average volt-amperes delivered to magnetic amplifier proper
$(\overline{VA})_L$ = average volt-amperes delivered to load
$(\overline{VA})_x$ = average core excitation volt-amperes
V_c = iron volume of magnetic core
V_T = total volume of wound core
V_w = winding volume of wound core
α = reset factor (Q_c^*/Q_R^*)
γ = dynamic range ratio $\Delta \bar{E}_L/(\bar{E}_L)_{co}$
η_c = control circuit efficiency $1/(1+R_S/r_c)$
η_L = load circuit efficiency $1/(1+r_x/R_L)$
ρ = resistivity of wire in given winding
τ_a = time for 63% response to step change in control voltage
ω_c = supply voltage angular frequency

Saturable Transformer

The saturable transformer has been introduced in Part I of this paper as the basic element of which all magnetic amplifiers are constructed. It has been found reasonable to represent the nonlinear excitation impedance of a transformer (utilizing the high-remanence cores typical of magnetic amplifiers) as an average resistance ($\bar{R}$). Thus, the equivalent circuit of such a transformer can be represented as in Fig. 1(A). Herein the star (*) is used to designate those parameters associated with the component saturable transformers as against those parameters associated with the magnetic amplifier as a whole. The a-c magnetization curve of Fig. 1(B) serves to illustrate the characteristics of a typical saturable transformer.

In Part I of this paper, the magnetic-design equations and performance parameters were expressed in terms of those parameters which are characteristic of the saturable transformer. Herein, it is intended to go one step further and relate the design equations to the physical and magnetic constants of the magnetic cores. In order to generalize this investigation the toroidal core is considered to be a solid uncased core which can be wound completely full.

The excitation impedance is given by

$$\bar{R} = (N_x^2)(\omega_c)[(\bar{B}_0/\bar{H}_0)(A_c/l)] \tag{1}$$

or more generally

$$\bar{R}_N/N^2 = (\omega_c)[(\bar{B}_0/\bar{H}_0)(A_c/l)] \tag{2}$$

The resistance of any winding (assuming a space factor of 0.4) is

$$r_N^* = (\rho)(N)(\bar{S}_N)\left(\frac{N}{0.4A_n}\right) \text{ or}$$

$$r_N^* = (\rho)(\bar{S}_N)\left(\frac{N^2}{0.4A_n}\right) \tag{3}$$

For simplicity, sector winding is assumed such that the mean length of turn of each winding ($\bar{S}_N$) is the same. The Q^* of a winding is defined as the ratio of its impedance to its winding resistance, thus

$$Q_N^* = \frac{(\omega_c)(B_0/H_0)(A_c/l)(0.4A_N)}{(\rho)(\bar{S}_N)} \tag{4}$$

or

$$Q_N^* = (K_1)(A_c/l)(A_N/\bar{S}_N) \tag{5}$$

where

$$K_1 = \left[\frac{(0.4)(\omega_c)}{(\rho)}\right](\bar{B}_0/\bar{H}_0) \tag{6}$$

The Q^* of a winding which completely fills the core window area A_w is

$$Q_T^* = (K_1)(A_c/l)(A_w/\bar{S}_0) \tag{7}$$

The minimum Q^* which a reset winding can have and still be capable of full control of the core corresponds to current-limited operation of the winding. Thus it has been shown (see Part I) that

$$Q_R^* = \frac{(\bar{H}_0 l)(Q_T^*)}{(0.4A_w)(D/k_R)} \tag{8}$$

or

$$Q_R^* = K_2(A_c/\bar{S}_0) \tag{9}$$

where

$$K_2 = \frac{(\omega_c)(\bar{B}_0)}{(\rho)(D/k_R)} \tag{10}$$

This minimum Q^* is a valuable normalizing factor. Therefore, the normalized control winding Q^* is

$$Q_c^*/Q_R^* = \alpha \tag{11}$$

and, since Q^* is proportional to winding area,

$$Q_T^* = Q_x^* + Q_c^* \tag{12}$$

Therefore,

$$(Q_T^*/Q_R^*) - \alpha = Q_x^*/Q_R^* \tag{13}$$

INTRODUCTION OF MAGNETIC-AMPLIFIER DESIGN EQUATIONS

In Part I of this paper it was shown that the output power $[(\overline{VA})_L]$ of a full-wave magnetic amplifier (see Fig. 2) normalized with respect to the core excitation power $[(\overline{VA})_x]$ is given by

$$\frac{(\overline{VA})_L}{(\overline{VA})_x} = Q_x\left[\frac{R_L/r_x}{(1+R_L/r_x)^2}\right] \tag{14}$$

and further that

$$\frac{(\overline{VA})_L}{(\overline{VA})_x} = Q_x^*\left[\frac{R_L/r_x^*}{(1+R_L/r_x^*)^2}\right] \tag{15}$$

The condition for maximum output power corresponds to current-limited operation of the load winding (N_x) and this may or may not be a matched condition, i.e., $R_L/r_x = 1$. The current capacity of the load winding limits the degree with which one can approach a matched condition in the same fashion that the current capacity of a reset winding limits the minimum usable Q.

In fact, the two conditions yield almost identical expression, thus

$$1+\min(R_L/r_x{}^*)=\frac{(\bar{H}_0 l)(Q_T{}^*)}{(0.4A_w)(D/k_x)(2)} \tag{16}$$

and from equation 8

$$\min(R_L/r_x{}^*)=(1/2)(k_x/k_R)(Q_R{}^*)-1 \tag{17}$$

For simplicity, the form factors of the load current (k_x) and of the reset current (k_R) are made identical (equal to 1.1) by assuming full-wave rectified reset current. Thus,

$$\min(R_L/r_x{}^*)=\frac{Q_R{}^*}{2}-1,\ k_x=k_R=1.1 \tag{18}$$

If the load winding (N_x) is operated at its rated current, then

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=Q_x{}^*\left\{\frac{\min(R_L/r_x{}^*)}{[1+\min(R_L/r_x)]^2}\right\} \tag{19}$$

or

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=Q_x{}^*\left[\frac{(Q_R{}^*-2)}{(Q_R{}^*)^2}\right](2) \tag{20}$$

Substitution of equation 13 into equation 20 yields

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=\left(\frac{Q_T{}^*}{Q_R{}^*}-\alpha\right)\left(1-\frac{2}{Q_R{}^*}\right)(2) \tag{21}$$

In order to simplify this investigation by removing frequency as a variable, it is necessary to assume that

$$Q_R{}^*\gg 2 \tag{22}$$

The fact that this is a reasonable assumption in the frequency range above approximately 1,000 radians per second can be verified by substituting typical values into equation 9. The general conclusions which are reached with the aid of this assumption fortunately are not invalidated in the low-frequency range in which Q_R is not much greater than one. It is interesting to note at this point that the most important single factor which determines whether or not the load can be matched is the supply frequency.

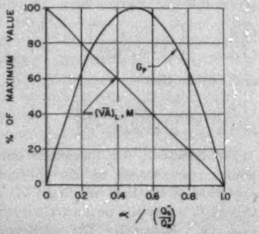


Fig. 3. Load power $[(\overline{VA})]_L/\omega_c$, figure of merit ($M/\omega_c$), and steady-state power gain (G_p/ω_c), normalized with respect to their maximum values, versus the normalized reset factor (α)

Application of expression 22 to equation 21 yields

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=\left(\frac{Q_T{}^*}{Q_R{}^*}-\alpha\right)(2),\ Q_R{}^*\gg 2 \tag{23}$$

It was also seen in Part I that the normalized input power is given by

$$\frac{(\overline{VA})_{IN}}{(\overline{VA})_x}=\frac{1}{Q_c}=\frac{2}{Q_c{}^*} \tag{24}$$

or

$$\frac{(\overline{VA})_{IN}}{(\overline{VA})_x}=\frac{2}{\alpha Q_R{}^*} \tag{25}$$

Therefore the steady-state power gain (G_p) can be determined from equations 23 and 25 as follows

$$G_p=\frac{(\overline{VA})_L}{(\overline{VA})_{IN}}=(\alpha Q_R{}^*)[(Q_T{}^*/Q_R{}^*)-\alpha] \tag{26}$$

Also, one definition of figure of merit M has been shown to be

$$M=\frac{G_p\eta_c}{\tau_a-T}=\left[\frac{(\overline{VA})_L}{(\overline{VA})_x}\right]\left(\frac{\omega_c}{\pi}\right) \tag{27}$$

Thus,

$$M=2[(Q_T{}^*/Q_R{}^*)-\alpha]\left(\frac{\omega_c}{\pi}\right) \tag{28}$$

It is interesting to note that the foregoing expression is substantially in agreement with the early work of Carleton and Horton.[4]

To remove supply frequency (ω_c) as a variable, the foregoing may be reduced as follows

$$\frac{(\overline{VA})_L}{(\overline{VA})_x}=\left[\frac{(\overline{VA})_L}{(\omega_c)(\bar{B}_0\bar{H}_0)(A_c l)}\right] =[(Q_T{}^*/Q_R{}^*)-\alpha](2) \tag{29}$$

or

$$(\overline{VA})_L/\omega_c=2[(\bar{B}_0\bar{H}_0)(V_c)][(Q_T{}^*/Q_R{}^*)-\alpha] \tag{30}$$

In addition,

$$G_p/\omega_c=(\alpha Q_R{}^*/\omega_c)(Q_T{}^*/Q_R{}^*-\alpha) \tag{31}$$

and from equation 9

$$Q_R{}^*/\omega_c=(K_2/\omega_c)(A_c/\bar{S}_0) \tag{32}$$

The application of equation 10 to the foregoing yields

$$Q_R{}^*/\omega_c=\left[\frac{(\bar{B}_0)}{(\rho)(D/k_R)}\right](A_c/\bar{S}_0) \tag{33}$$

or

$$Q_R{}^*/\omega_c=K_3(A_c/\bar{S}_0) \tag{34}$$

where

$$K_3=\frac{\bar{B}_0}{(\rho)(D/k_R)} \tag{35}$$

Clearly from equation 28

$$M/\omega_c=2[(Q_T{}^*/Q_R{}^*)-\alpha](1/\pi) \tag{36}$$

The factor $(Q_T{}^*/Q_R{}^*)$ can be seen to be inherently independent of frequency by combining equations 7 and 9.

Thus

$$(Q_T{}^*/Q_R{}^*)=(K_1/K_2)(A_w/l) =(0.4A_w)(D/k_R)/(\bar{H}_0 l) \tag{37}$$

For purposes of this investigation, the change in $\bar{H}_0$ due to frequency is assumed negligible. In addition, since the change in $\bar{H}_0$ caused by changing ID/OD is normally no greater than the differences between samples of the same material, this factor is neglected.

At this point is is necessary to investigate the effect of α so as to be able to remove it as a variable. The ultimate objective of course is to reduce the expressions for output power, power gain, etc., to functions only of the physical dimensions and/or dimensional ratios of the toroidal core. In Fig. 3 the normalized output power, figure of merit, and steady-state power gain are plotted versus the normalized α. Clearly the condition for maximum $[(\overline{VA})_L/\omega_c]$ and maximum (M/ω_c) obtains when α is zero. Note that the maximum power-gain condition corresponds to $Q_c{}^*$ equal to one half of $Q_T{}^*$, i.e., the control winding occupies one half of the window area. Clearly a compromise value of α is necessary if it is to be eliminated as a variable. Quite naturally this value has been taken as

$$\alpha=(0.25)(Q_T{}^*/Q_R{}^*)$$

Thus, at this point, the important parameters $(\overline{VA}_L/\omega_c)$, (M/ω_c), and (G_p/ω_c) each has a value corresponding to 75% of the maximum attainable; i.e.,

$$(\overline{VA})_L/\omega_c=(0.75)[(\bar{B}_0\bar{H}_0)(V_c)(2)\times (Q_T{}^*/Q_R{}^*)],\ \alpha=(Q_T{}^*/Q_R{}^*)/4 \tag{38}$$

and

$$G_p/\omega_c=(0.25Q_T{}^*/Q_R{}^*)(Q_R{}^*/\omega_c)\times (0.75Q_T{}^*/Q_R{}^*)$$

or

$$G_p/\omega_c=(0.1875)(Q_R{}^*/\omega_c)(Q_T{}^*/Q_R{}^*)^2,\ \alpha=(Q_T{}^*/Q_R{}^*)/4 \tag{39}$$

Also

$$M/\omega_c=(0.75)[2(Q_T{}^*/Q_R{}^*)(1/\pi)],\ \alpha=(Q_T{}^*/Q_R{}^*)/4 \tag{40}$$

The subsequent work is such that any other finite value of α could have been selected with the same end result.

In the foregoing, the important parameters which describe magnetic-amplifier performance have been reduced to functions of two variables, namely $(Q_T{}^*/Q_R{}^*)$

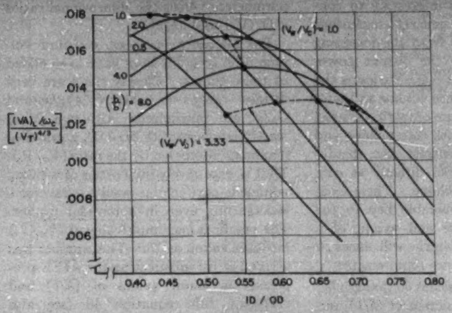


Fig. 4. Load power (normalized with respect to carrier frequency and total transformer volume) versus toroidal-core dimensional ratios h/b and ID/OD

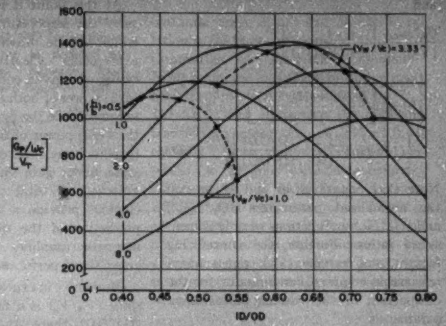


Fig. 5. Steady-state power gain (normalized with respect to carrier frequency and total transformer volume) versus toroidal-core dimensional ratios h/b and ID/OD

and (Q_R^*/ω_c). These two variables in turn are functions only of the physical dimensions of the core, i.e.,

$$(Q_T^*/Q_R^*) = (K_1/K_2)(A_w/l) \tag{37}$$

and

$$Q_R^*/\omega_c = K_2(A_c/\bar{S}_0) \tag{34}$$

It remains to reduce further (A_w/l) and $(A_c/\bar{S}_0)$ to more basic dimensions and/or dimensional ratios. It has been found that the most meaningful of these latter quantities are:

1. (ID/OD), the ratio of core inside diameter to outside diameter.
2. (h/b), the ratio of core height to build-up.
3. V_c, the volume of the core.
4. V_w/V_c, the ratio of winding volume to core volume.
5. V_T, the total volume of the wound core

Clearly, if an optimum core shape exists, this optimum should be independent of the physical size of the core. That is, those parameters which are descriptive of magnetic-amplifier performance should be reducible to functions of only dimensional ratios such that the absolute values of the dimensions merely act as scale factors to change the physical size but not the shape of the core. In line with this reasoning (Q_T^*/Q_R^*) and (Q_R^*/ω_c) are next normalized with respect to size, i.e., V_c and V_T, such that they become functions only of dimensional ratios. Thus using equation 27 and noting that

$$\frac{A_w}{l} = \left[\frac{V_c}{(\pi)(h/b)\left(\frac{1}{\text{ID/OD}}-1\right)^2\left(\frac{1}{\text{ID/OD}}+1\right)^4}\right]^{1/3} \tag{41}$$

it can be seen that

$$\frac{(Q_T^*/Q_R^*)}{(V_c)^{1/3}} = \frac{(K_1)}{(K_2)} \times \left[\frac{1}{(\pi)(h/b)\left(\frac{1}{\text{ID/OD}}-1\right)^2\left(\frac{1}{\text{ID/OD}}+1\right)^4}\right]^{1/3} \tag{42}$$

It has been found that the mean length of turn $(\bar{S}_0)$ of a winding which completely fills the window of a toroidal core is given with good approximation by

$$\bar{S}_0 = 2h \times \left\{1+\frac{(0.98)[(\text{ID/OD})+0.306]}{[1-(\text{ID/OD})](h/b)}\right\}, \quad \text{ID/OD} \leq 0.9 \tag{43}$$

(This expression was derived from a more general formulation of the mean length of turn of any winding on a toroid which was published in a Boeing Airplane Company internal memorandum by Louis Miovski.) Therefore, the use of equation 43 permits equation 34 to be reduced to the following

$$\frac{Q_R^*/\omega_c}{(V_c)^{1/3}} = \left\{\frac{(0.5)(K_2)}{(0.98)\left[\frac{0.306+(\text{ID/OD})}{1-(\text{ID/OD})}\right]+(h/b)}\right\} \times \left\{(1/\pi)(h/b)^2\left[\frac{1-(\text{ID/OD})}{1+(\text{ID/OD})}\right]\right\}^{1/3} \tag{44}$$

Since

$$V_T = V_c[1+(V_w/V_c)] \tag{45}$$

equations 42 and 44 can also be put in the form

$$\frac{Q_T^*/Q_R^*}{(V_T)^{1/3}} = \left(\frac{K_1}{K_2}\right)\left[\frac{1}{1+(V_w/V_c)}\right] \times \left[\frac{1}{(\pi)(h/b)\left(\frac{1}{\text{ID/OD}}-1\right)^2\left(\frac{1}{\text{ID/OD}}+1\right)^4}\right]^{1/3} \tag{46}$$

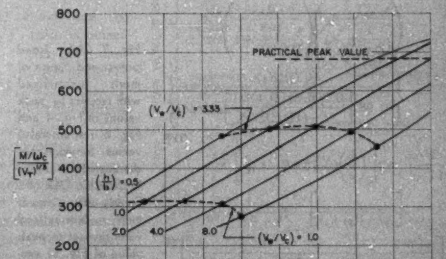


Fig. 6. Figure of merit (normalized with respect to carrier frequency and total transformer volume) versus toroidal-core dimensional ratios h/b and ID/OD

and

$$\frac{Q_R^*/\omega_c}{(V_T)^{1/3}}=\left[\frac{1}{1+(V_w/V_c)}\right]^{1/3}\times$$

$$\left\{\frac{(0.5)(K_3)}{(0.98)\left[\frac{0.306+(\mathrm{ID/OD})}{1-(\mathrm{ID/OD})}\right]+(h/b)}\right\}\times$$

$$\left\{(1/\pi)(h/b)^2\left[\frac{1-(\mathrm{ID/OD})}{1+(\mathrm{ID/OD})}\right]\right\}^{1/3} \tag{47}$$

Note that equations 46 and 47 present the normalized parameters (Q_T^*/Q_R^*) and (Q_R^*/ω_c) as functions of only dimensional ratios. Further, the normalizing factor, total volume (V_T), represents a reasonable common denominator for the comparison of the various performance parameters.

After some algebraic manipulation it is found that

$$\frac{V_w}{V_c}=\frac{(A_w\bar{S}_0)}{(bhl)}=2\times$$

$$\left\{\frac{(0.98)\left[\frac{0.306+(\mathrm{ID/OD})}{1-(\mathrm{ID/OD})}\right]+(h/b)}{\left(\frac{1}{\mathrm{ID/OD}}+1\right)\left(\frac{1}{\mathrm{ID/OD}}-1\right)(h/b)}\right\} \tag{48}$$

and

$$V_c=(h^3)\left\{\frac{(\pi)[1+(\mathrm{ID/OD})]}{(h/b)^2[1-(\mathrm{ID/OD}]}\right\} \tag{49}$$

Using equations 45, 48, and 49 and the expressions in equations 46 and 47, equations 38 through 40 can be reduced to functions of dimensional ratios only Thus,

$$\frac{(\overline{VA})_L/\omega_c}{(V_T)^{4/3}}=2[(0.75)(\bar{B}_0\bar{H}_0)(K_1/K_2)]\times$$

$$\left[\frac{1}{1+(V_w/V_c)}\right]^{4/3}\times$$

$$\left[\frac{1}{(\pi)(h/b)\left(\frac{1}{\mathrm{ID/OD}}-1\right)^2\left(\frac{1}{\mathrm{ID/OD}}+1\right)^4}\right]^{1/3} \tag{50}$$

$$\frac{G_p/\omega_c}{V_T}=[(0\ 06)(K_1/K_2)^2]\left[\frac{1}{1+(V_wV_c)}\right]\times$$

$$\left(\frac{1}{\frac{1}{\mathrm{ID/OD}}-1}\right)\left(\frac{1}{\frac{1}{\mathrm{ID/OD}}+1}\right)^3\times$$

$$\left\{\frac{(0.5)(K_3)}{(0.98)\left[\frac{0.306+(\mathrm{ID/OD})}{1-(\mathrm{ID/OD})}\right]+(h/b)}\right\} \tag{51}$$

$$\frac{M/\omega_c}{(V_T)^{1/3}}=2\left[\left(\frac{1}{\pi}\right)(0.75)(K_1/K_2)\right]\times$$

$$\left[\frac{1}{1+(V_w/V_c)}\right]^{1/3}\times$$

$$\left[\frac{1}{(\pi)(h/b)\left(\frac{1}{\mathrm{ID/OD}}-1\right)^2\left(\frac{1}{\mathrm{ID/OD}}+1\right)^4}\right]^{1/3} \tag{52}$$

At this point it is interesting to consider the important general conclusions which can be drawn from the foregoing expressions. Clearly the output power rating of a given core shape varies as the (4/3) power of both core volume and total volume, while the figure of merit varies as the (1/3) power of both core volume and total volume. In contrast, the steady-state power gain varies directly as core volume and total volume. These conclusions provide a valuable key to the prediction of the size and weight of a magnetic amplifier which will meet a given set of performance requirements.

Obviously, in expressions 50 through 52, since (V_w/V_c) is a function of (h/b) and (ID/OD), there are only two variables, namely, (h/b) and (ID/OD). Note that (K_1/K_2) and K_3, which are dependent on the kind of core material, allowable current density, etc., are merely scale factors which can have no effect on the shape of the curves for the foregoing three expressions. Therefore, for simplicity, typical values for K_1/K_2, K_3, and $\bar{B}_0\bar{H}_0$ have been utilized in the following (see the Appendix). Figs. 4 through 6 respectively present plots of the three performance parameters versus the dimensional ratios (ID/OD) and (h/b).

For reference purposes, constant volume ratio (V_w/V_c) lines have been added to these curves. Clearly there are very definite ranges in which $(\overline{VA})_L/\omega_c$ and G_p/ω_c are maximized. Although M/ω_c merely increases with ID/OD, a practical maximum is imposed by the ratio (V_w/V_c). That is, excessively high ratios of winding volume to core colume results in excessive leakage flux even in a toroidal reactor. The practical maximum value of (V_w/V_c) has been taken as 15. This number was arrived at by substituting the AIEE proposed maximum values of (h/b) and (ID/OD) into equation 48 (see also Fig. 9).

The peak values can be determined from the three curves to be:

$$\text{Peak}\left|\frac{(VA)_L/\omega_c}{(V_T)^{4/3}}\right|=0.018$$

$$\text{Peak}\left|\frac{G_p/\omega_c}{(V_T)}\right|=1,420$$

$$\text{Peak}\left|\frac{M/\omega_c}{(V_T)^{1/3}}\right|=682$$

If each of the curves (Figs. 4 through 6)

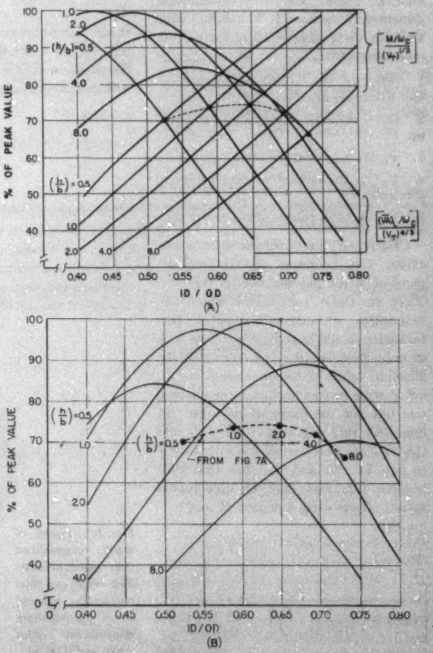


Fig. 7 (A). Load power and figure of merit (normalized with respect to peak values of Fig. 4 and Fig. 6 respectively) versus toroidal-core dimensional ratios h/b and ID/OD. (B). Steady-state power gain (normalized with respect to peak value of Fig. 5) versus the toroidal core dimensional ratios h/b and ID/OD

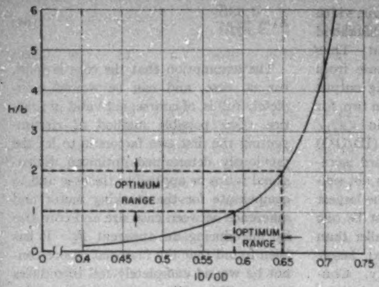


Fig. 8. Relationship between h/b and ID/OD for optimum toroidal core as dictated by compromise between load power and figure of merit

is normalized with respect to the appropriate peak value, the plots in Fig. 7 result. These plots permit the selection of those dimensional ratios which represent the best compromise between all three key performance parameters. If $(\overline{VA})_L/\omega_c$ and M/ω_c are weighed as being of equal importance, then the intersections of the appropriate normalized curves, Fig. 7(A), represent a compromise between these two factors. Fortunately these intersections, Fig. 7(B), occur very near the peak values for the corresponding G_p/ω_c. Therefore, under compromise conditions, both $(\overline{VA})_L/\omega_c$ and M/ω_c are reduced to approximately 70% to 75% of their peak values, and G_p/ω_c is approximately 90% of the peak value in the ID/OD range between approximately 0.55 and 0.70. The analytical solution for the intersection of the $(\overline{VA})_L/\omega_c$ and the M/ω_c curves corresponds to a constant (V_T/V_c) ratio of 3.33 which (using equation 48 can be reduced to

$$(h/b) = \left[\frac{(0.98)[0.306+(\mathrm{ID/OD})]}{[1-(\mathrm{ID/OD})]\left[(1.667)\times\left(\frac{1}{\mathrm{ID/OD}}+1\right)\left(\frac{1}{\mathrm{ID/OD}}-1\right)-1\right]}\right] \tag{53}$$

This expression is plotted in Fig. 8 and clearly illustrates that a reasonable compromise between output power and figure of merit is impossible for values of ID/OD greater than 0.7 if the practical maximum value of h/b is taken as four. The relationship in equation 53 is the key to the subsequent selection of optimum standard core sizes.

The ratio of winding volume to core volume (V_w/V_c) is obviously a very important parameter in view of the fact that the locus of the compromise values of (ID/OD) and (h/b) corresponds to a constant volume ratio of 3.33. It is also interesting to note that the locus of the peak values of normalized output power (Fig. 4) corresponds to a constant (V_w/V_c) ratio of one. Although in the magnetic amplifier it is undesirable to optimize the load power at the expense of power gain and figure of merit, such an optimization is possible in a simple transformer. Thus a (V_w/V_c) ratio of one is ideal for toroidal transformers and the relationship between (h/b) and (ID/OD) would properly be

$$h/b = \left[\frac{(0.98)[0.306+(\mathrm{ID/OD})]}{[1-(\mathrm{ID/OD})]\left[(0.5)\times\left(\frac{1}{\mathrm{ID/OD}}+1\right)\left(\frac{1}{\mathrm{ID/OD}}-1\right)-1\right]}\right] \tag{54}$$

In the opinion of the authors, the 25% reduction in output power and figure of merit which obtains under compromise conditions is not sufficient to warrant the selection of two classes of core sizes, i.e., high power versus high gain, for magnetic-amplifier application. However, equation 54 could be used as the basis for the selection of optimum toroidal transformer cores.

A study of Fig. 7(A) reveals that, as long as the proper V_w/V_c ratio is maintained, both (h/b) and (ID/OD) can vary over a wide range with but little effect on the values of $(\overline{VA})_L/\omega_c$ and M/ω_c. However, very definite optimum ranges for these parameters are dictated, Fig. 7(B), by the steady-state power gain G_p/ω_c. Note that in the h/b range between one and two, G_p/ω_c is approximately 95% of the peak value but that it falls off rapidly for large and for smaller values of h/b. In view of this and also in view of the practical limitations on the value of h/b, the h/b range between one and two has been selected as optimum. Fig. 9 serves to clarify the location of the optimum range of h/b and ID/OD in relation to the limits proposed by the AIEE Committee on Magnetic Amplifiers.[8] In addition, the location of optimum magnetic-amplifier cores (determined by $V_w/V_c=3.33$) with respect to the optimum transformer cores, i.e., $V_w/V_c=1.0$, is clarified. Note that the common criterion for the good toroidal transformer core, namely a unity ratio between window area and core area (A_w/A_c), corresponds very roughly to a V_w/V_c of one. However, in the opinion of the authors, the latter parameter is a better measure of the performance to be expected from a toroidal transformer.

Proposed Standard Core Sizes

The utilization of the foregoing information in a logical manner can yield a list of optimum shaped cores which effectively covers the practical range of sizes. The output power capacity of the core $(\overline{VA})_L/\omega_c$ has been selected as the parameter which determines the size of the steps between core sizes. After the fashion of wire sizes, it has been decided to arrange standard cores in the order of increasing power on a logarithmic scale. Therefore, it is proposed to have ten cores for each order of magnitude of output power rating, with the even-numbered cores being preferred sizes. It is further intended to utilize the proposed AIEE standard tape widths. The fact that it is impossible to also utilize only the proposed standard inside diameters can be verified by noting that

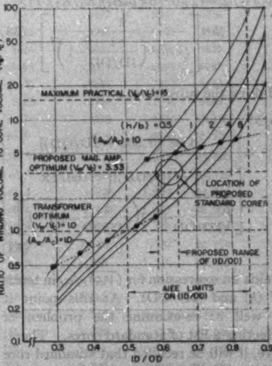


Fig. 9. Ratio of winding volume to core volume versus toroidal-core dimensional ratios h/b and ID/OD, illustrating location of proposed optimum standard toroidal cores

$$\frac{\mathrm{ID}}{h}=\left[\frac{2}{(h/b)\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)}\right] \tag{55}$$

Clearly if both ID and h are only standard sizes, the relationship between (h/b) and (ID/OD) is predetermined and the optimum condition described by equation 53 cannot be utilized. Therefore, if either (ID) or (h) is to be standardized, the other must be allowed to have nonstandard values. In the opinion of the authors, standard values of tape width (h) are very desirable while nonstandard ID represent only a minor inconvenience. Of course, this opinion could lead to a philosophical argument about the relative merits of complete standardization on a minimum number of standard dimensions versus the resultant compromised performance of the toroidal core. It is felt that at this stage in the growth of magnetic amplifiers, an unreasonably restrictive standardization is unrealistic. When one compares the added complexity introduced by nonstandard ID to the present lack of standardization in the vacuum-tube field, the former factor seems quite insignificant.

To evaluate $(\overline{VA})_L/\omega_c$, equation 38 can be reduced (by using equation 37) to

$$(\overline{VA})_L/\omega_c=2[(0.75)(\overline{B}_0\overline{H}_0)(K_1/K_2)]\times(A_cA_w) \tag{56}$$

where

$$(A_cA_w)=(h^4)\left[\frac{\pi}{(h/b)^3\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)^2}\right] \tag{57}$$

From this, it can be seen that

$$(\overline{VA})_L/\omega_c=(2)(0.75)(B_0H_0)(K_1/K_2)(\pi)\times(h^4)\left[\frac{1}{(h/b)^3\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)^2}\right] \tag{58}$$

Making the substitution

$$(h/b)=\left[\frac{(0.98)[0.306+(\mathrm{ID/OD})]}{[1-(\mathrm{ID/OD})]\left[(1.667)\times\left(\dfrac{1}{\mathrm{ID/OD}}+1\right)\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)-1\right]}\right] \tag{53}$$

yields an expression for $(\overline{VA})_L/\omega_c$ in terms of (h) and (ID/OD). At this point it is well to re-examine the problem of selecting a list of standard cores. Therefore, it will be recalled that standard core heights ($^1/_8$, $^1/_4$ inch, etc.) must be utilized and (ID/OD) must be kept in the optimum range which falls approximately between 0.575 and 0.675 (see Fig. 9). Further, it has been arbitrarily decided that the largest core of a given tape width should have an h/b in the neighborhood of one, i.e., square cross section. Thus (h/b) will be allowed to increase from one (in the process of changing output power) rather than decrease from two, for example. Further, the increase in $(\overline{VA})_L/\omega_c$ which obtains for a constant (ID/OD) when (h) is increased in standard increments ($^1/_8$, $^3/_{16}$, $^1/_4$ inch, etc.) is not uniform, therefore, the fact that the largest core of a given height (h) must be one increment of output power smaller than the smallest core of the next larger height, leads to some difficulty. Consideration of all these factors makes the analytical solution of equation 58 rather tedious. It has been found that a graphical solution is more straightforward.

Inherent in equations 50 through 52 is the assumption that the control winding occupies one fourth of the window area, i.e., $\alpha=(Q_T^*/Q_R^*)/4$. Since the actual value of α merely acts as a scale factor on the absolute value of output power, etc., it has no effect on the previous optimization process based on normalized values (Fig. 7). However, it is desirable, in the tabulation of core parameters, to itemize the maximum attainable values of $(\overline{VA})_L/\omega_c$, M/ω_c, and G_p/ω_c. Therefore, since each of these quantities has 75% of its maximum value in equations 50 through 52 it is a simple process to determine these maximums. In addition, it is desirable to so generalize the tabulation as to make it independent of the characteristics of the core material, and the allowable current density, etc. Therefore, equations 50 through 52 can be modified to yield

$$\max[(\overline{VA})_L/\omega_c](K_4)=(V_c)^{4/3}\times\left[\frac{1}{(h/b)\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)^2\left(\dfrac{1}{\mathrm{ID/OD}}+1\right)^4}\right]^{1/3} \tag{59}$$

$$K_4=\frac{2.03}{(D)(\overline{B}_0)} \tag{60}$$

$$\max(G_p/\omega_c)(K_5)=(V_c)\times\left(\frac{1}{\dfrac{1}{\mathrm{ID/OD}}-1}\right)\left(\frac{1}{\dfrac{1}{\mathrm{ID/OD}}+1}\right)^2\times\left\{(0.98)\left(\frac{0.306+(\mathrm{ID/OD})}{1-(\mathrm{ID/OD})}\right)+(h/b)\right\} \tag{61}$$

$$K_5=\left(\frac{176\rho}{D}\right)\left[\frac{\overline{H}_0}{(\overline{B}/\overline{H}_0)}\right] \tag{62}$$

$$\max(M/\omega_c)(K_6)=(V_c)^{1/3}\times\left[\frac{1}{(h/b)\left(\dfrac{1}{\mathrm{ID/OD}}-1\right)^2\left(\dfrac{1}{\mathrm{ID/OD}}+1\right)^4}\right]^{1/3} \tag{63}$$

$$K_6=\frac{6.38\overline{H}_0}{D} \tag{64}$$

The assumption that the core is solid, has no case, and can be wound completely full is, of course, not valid in practice. One possible method of circumventing the first two factors is to let the previously determined optimum dimensional ratios be applied to the case and to compensate for the stacking factor and difference between case size and core size by introducing an apparent $\overline{B}_0$. It has been found that the fact that a core cannot be wound completely full introduces a nonuniform discrepency between the idealized and the practical results. This latter factor is virtually impossible to generalize because of the differences between the capabilities of different winding machines. It has been arbitrarily decided to select standard cores on the basis of the idealized solid core with no case and subsequently to tabulate the resultant practical values of the various parameters of interest.

It is interesting to note that equation 56 indicates that the power rating of a core is proportional to the product of core area and window area as has been previously indicated in the report of the AIEE Committee on Magnetic Amplifiers.[3] This one factor (A_cA_w) of course does not yield sufficient information to permit the selection of optimum core configurations.

The proposed list of standard cores appears in Table I. The tabulated 63 cores cover more than six orders of magnitude of normalized output power $(\overline{VA})_L/\omega_c$.

After the fashion of wire tables, a logarithmic scale has been used such that there are ten cores per decade of power. Therefore, an increase of three core sizes doubles the power rating, an increase of ten core sizes increases the rating by a factor of ten, etc. This uniformity in the core sizes should prove to be a great aid in the simplification of magnetic-amplifier design. In addition to the uniform increments of power between core sizes, the other parameters, power gain, and figure of merit also vary in a uniform manner. For example, since the output power varies as the (4/3) power of total volume and the figure of merit varies as the (1/3) power of total volume, it can be seen that figure of merit varies as the (1/4) power of output power. Thus the figure of merit increases by $(2)^{1/4}$ or a factor of 1.19 for each increase of three core sizes. Likewise the power gain increases by $(2)^{3/4}$ or a factor of 1.68 for each three core sizes. It is proposed, again after the

TABLE I PROPOSED STANDARD CORES

CORE NO.	ID IN	OD IN	H IN	H/B	ID/OD	ℓ IN	A_w IN^2	A_c IN^2	A_cA_w IN^4	V_c IN^3	$K_4[(VA)_L/\omega_c]$	$K_5[G_p/\omega_c]$	$K_6[M/\omega_c]$	$\frac{1}{2}(VA)_L$ (FOR HYMU-80 @ 1 KC)
1	0.227	0.351	0.1250	2.02	0.6473	0.908	0.040	0.0077	0.0003	0.0070	0.00046	0.00018	0.06537	0.482
2	.243	.380	.1250	1.82	.6395	.979	.046	.0086	.0004	.0084	.00085	.00023	.06945	.609
3	.260	.411	.1250	1.65	.6316	1.054	.053	.0095	.0005	.0100	.00074	.00024	.07325	.768
4	.277	.445	.1250	1.49	.6235	1.134	.060	.0105	.0006	.0119	.00092	.00028	.07802	.968
5	.296	.482	.1250	1.35	.6151	1.222	.069	.0116	.0008	.0142	.00117	.00034	.08262	1.215
6	.316	.520	.1250	1.22	.6067	1.313	.078	.0128	.0010	.0168	.00146	.00040	.08730	1.523
7	.337	.563	.1250	1.11	.5979	1.414	.089	.0142	.0013	.0200	.00185	.00047	.09230	1.920
8	.359	.609	.1250	1.00	.5890	1.521	.101	.0156	.0016	.0238	.00232	.00055	.09744	2.412
9	.364	.569	.1875	1.83	.6397	1.466	.104	.0192	.0020	.0282	.00293	.00068	.1040	3.050
10	.388	.614	.1875	1.66	.6320	1.575	.118	.0212	.0025	.0334	.00368	.00080	.1102	3.830
11	.415	.665	.1875	1.50	.6239	1.696	.135	.0234	.0032	.0397	.00464	.00095	.1167	4.820
12	.443	.720	.1875	1.36	.6155	1.826	.154	.0250	.0040	.0474	.00585	.00113	.1236	6.087
13	.473	.779	.1875	1.22	.6069	1.967	.176	.0287	.0050	.0565	.00735	.00134	.1308	7.680
14	.504	.844	.1875	1.11	.5981	2.118	.200	.0318	.0063	.0673	.00925	.00158	.1383	9.680
15	.537	.912	.1875	1.00	.5893	2.276	.227	.0351	.0080	.0799	.0117	.00186	.1459	12.15
16	.554	.888	.2500	1.50	.6237	2.265	.241	.0418	.0100	.0946	.0146	.00227	.1558	15.23
17	.591	.960	.2500	1.36	.6155	2.435	.274	.0461	.0126	.1123	.0185	.00266	.1648	19.20
18	.630	1.038	.2500	1.22	.6070	2.620	.312	.0510	.0159	.1336	.0232	.00317	.1743	24.12
19	.672	1.124	.2500	1.11	.5982	2.821	.355	.0564	.0200	.1592	.0293	.00375	.1842	30.50
20	.716	1.215	.2500	1.00	.5894	3.032	.402	.0623	.0251	.1890	.0368	.00441	.1944	38.30
21	.726	1.134	.3750	1.84	.6401	2.920	.414	.0765	.0316	.2234	.0464	.00535	.2074	48.20
22	.775	1.227	.3750	1.66	.6322	3.145	.472	.0846	.0399	.2661	.0585	.00638	.2200	60.87
23	.828	1.327	.3750	1.50	.6241	3.385	.539	.0935	.0502	.3167	.0735	.00759	.2331	76.80
24	.885	1.437	.3750	1.36	.6157	3.646	.615	.1035	.0633	.3775	.0925	.00902	.2469	96.80
25	.943	1.553	.3750	1.23	.6075	3.920	.698	.1143	.0800	.4482	.117	.01065	.2609	121.5
26	1.005	1.680	.3750	1.11	.5986	4.217	.794	.1264	.1004	.5331	.146	.01257	.2757	152.3
27	1.072	1.817	.3750	1.00	.5897	4.538	.902	.1398	.1261	.6343	.185	.01479	.2911	129.0
28	1.102	1.766	.5000	1.51	.6253	4.506	.955	.1659	.159	.7474	.232	.01793	.3103	241
29	1.178	1.912	.5000	1.36	.6159	4.854	1.089	.1837	.2000	.8912	.293	.02131	.3287	305
30	1.255	2.066	.5000	1.23	.6075	5.218	1.238	.2028	.2510	1.058	.368	.02505	.3474	383
31	1.338	2.235	.5000	1.11	.5988	5.614	1.407	.2242	.3155	1.250	.464	.02968	.3671	482
32	1.429	2.423	.5000	1.00	.5897	6.050	1.608	.2485	.399	1.504	.585	.03507	.3861	609
33	1.448	2.260	.750	1.845	.6404	5.82	1.646	.305	.502	1.78	.735	.0425	.4139	768
34	1.546	2.444	.750	1.671	.6326	6.27	1.877	.377	.633	2.11	.925	.0564	.4387	968
35	1.651	2.644	.750	1.511	.6245	6.75	2.141	.372	.800	2.51	1.17	.0625	.4648	1215
36	1.762	2.860	.750	1.366	.6162	7.26	2.440	.412	1.004	2.99	1.46	.0715	.4921	1523
37	1.879	3.091	.750	1.237	.6078	7.81	2.772	.455	1.26	3.55	1.85	.0844	.5201	1920
38	2.005	3.347	.750	1.118	.5990	8.41	3.157	.503	1.59	4.23	2.32	.0998	.5500	2412
39	2.138	3.625	.750	1.009	.5900	9.05	3.592	.557	2.00	5.04	2.93	.1177	.5811	3050
40	2.200	3.522	1.000	1.513	.6246	8.99	3.800	.661	2.51	5.94	3.68	.1425	.6193	3830
41	2.348	3.810	1.000	1.368	.6163	9.67	4.330	.731	3.155	7.07	4.64	.1691	.6556	4820
42	2.507	4.125	1.000	1.236	.6077	10.4	4.930	.809	3.99	8.43	5.85	.2004	.6939	6087
43	2.673	4.463	1.000	1.118	.5990	11.2	5.612	.895	5.022	10.03	7.35	.2366	.7333	7680
44	2.851	4.833	1.000	1.009	.5900	12.1	6.386	.991	6.33	11.96	9.25	.2790	.7748	9680
45	2.868	4.507	1.500	1.852	.6407	11.6	6.550	1.215	8.00	14.11	11.7	.3379	.8258	12,150
46	3.062	4.868	1.500	1.679	.6330	12.5	7.430	1.340	10.04	16.73	14.6	.4016	.8748	15,230
47	3.292	5.268	1.500	1.518	.6249	13.4	8.510	1.482	12.61	19.92	18.5	.4779	.9271	19,200
48	3.517	5.704	1.500	1.371	.6165	14.5	9.713	1.641	15.93	23.76	23.2	.5684	.9822	24,120
49	3.752	6.171	1.500	1.240	.6080	15.6	11.056	1.814	20.0	28.28	29.3	.6726	1.039	30,500
50	3.998	6.670	1.500	1.123	.5994	16.8	12.556	2.004	25.1	33.58	36.8	.7925	1.097	38,300
51	4.262	7.218	1.500	1.015	.5905	18.0	14.267	2.217	31.55	39.97	46.4	.9333	1.158	48,200
52	4.392	7.030	2.000	1.516	.6248	17.9	15.154	2.630	39.0	47.33	58.5	1.135	1.237	60,870
53	4.693	7.613	2.000	1.370	.6164	19.3	17.295	2.920	50.22	56.45	73.5	1.350	1.310	76,800
54	5.003	8.228	2.000	1.240	.6080	20.8	19.656	3.225	63.30	67.03	92.5	1.594	1.385	96,800
55	5.335	8.902	2.000	1.121	.5993	22.4	22.352	3.567	80.00	79.76	117	1.882	1.464	121,500
56	5.683	9.623	2.000	1.015	.5905	24.0	25.363	3.941	100.4	94.78	146	2.212	1.545	152,300
57	5.756	8.978	3.000	1.862	.6411	23.1	26.021	4.833	126.1	111	185	2.678	1.647	192,000
58	6.138	9.689	3.000	1.689	.6335	24.9	29.589	5.326	159.3	132	232	3.177	1.743	241,200
59	6.568	10.505	3.000	1.524	.6252	26.8	33.879	5.906	200.0	158	293	3.799	2.850	305,000
60	7.006	11.355	3.000	1.379	.6170	28.8	38.554	6.524	251.0	188	368	4.501	1.958	383,000
61	7.476	12.286	3.000	1.247	.6085	31.0	43.894	7.215	315.5	224	464	5.329	2.071	482,000
62	7.973	13.293	3.000	1.128	.5998	33.4	49.928	7.980	399	266	585	6.292	2.189	608,700
63	8.506	14.397	3.000	1.018	.5908	36.0	56.825	8.837	502.2	318	735	7.426	2.313	768,000

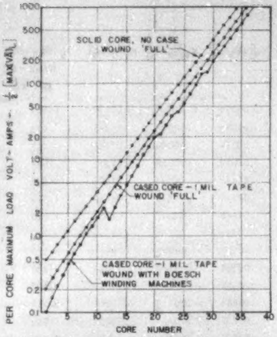


Fig. 10. Idealized maximum output power and practical maximum output power for proposed standard toroidal cores

fashion of wire tables, that the even-numbered cores be preferred sizes.

For illustration purposes, max $[(\overline{VA})_L]$ has also been tabulated for each core in Table I at a frequency of 1,000 cycles per second using typical Hymu-80 core material and the previously mentioned allowable current density. The difference between this theoretical limit and the output power attainable from a practical core (with core box, stacking factor, and minimum hole ID introduced) is illustrated by Fig. 10. Note that if the practical core can be wound completely full, i.e., no hole, a smooth curve results, but that the introduction of the practical limitations of typical winding machines produces a nonuniform difference between theoretical and practical results.

Applicability to Half-Wave Magnetic Amplifiers

The fact that the foregoing is also valid for half-wave circuits can be seen by noting that the load circuit, and therefore output power, considerations are the same. In addition, the steady-state power gain and the figure of merit differ only by a constant (π/ω_c) such that the previous compromise between output power and figure of merit is still valid.

Conclusions

An analytical method has been presented whereby the effect of toroidal-core dimensional ratios (h/b and ID/OD) on three major magnetic-amplifier performance parameters can be evaluated. On the basis of this evaluation, the optimum functional relationship between these dimensional ratios, in addition to the optimum range of values for the ratios, has been determined. Utilizing the foregoing information, a list of proposed standard cores which uniformly covers the practical range of sizes has been presented.

Appendix. Typical Values for the Constants (K_1/K_2),(K_3), and $(\bar{B}_0\bar{H}_0)$

For illustration purposes a typical toroid utilizing Hymu-80 core material has been selected. It is assumed that:

1. The winding temperature is 25 degrees centigrade such that $\rho=6.79\times10^{-7}$ ohm-inch.
2. The maximum allowable current density is $D=2{,}546$ amperes per square inch (or 2 milliamperes per circular mil).
3. The form factor of the reset current is $k_R=1.0$.
4. The average flux density is $\bar{B}_0=2.645\times10^{-4}$ webers per square inch (or 0.41 webers per square meter).
5. The average magnetizing force is $\bar{H}_0=0.1372$ ampere-turn per inch (or 5.4 ampere-turns per meter).

Therefore,

$$|K_1/K_2|=\frac{0.4(D/k_R)}{\bar{H}_0}=7{,}430$$

and

$$|K_3|=\frac{\bar{B}_0}{\rho(D/k_R)}=0.153$$

Also,

$$|\bar{B}_0\bar{H}_0|=3.63\times10^{-5}$$

References

1. Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques, Robert M. Hubbard. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

2. An Analysis of Optimum Core Configuration for Magnetic Amplifiers Using a Simplified Approach, B. D. Bedford, C. H. Willis, G. C. Dodson. *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 62–70.

3. Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores, AIEE Committee Report. *Ibid.*, vol. 76, pt. I, May 1957, pp. 135–37.

4. The Figure of Merit of Magnetic Amplifiers, J. T. Carleton, W. F. Horton. *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 239–45.

◆

Discussion

Edward Korn (Aerophysics Development Corporation): While the authors' statement that "the change in H_0 caused by changing ID/OD is normally no greater than the differences between samples of the same material" may be true, it does not necessarily follow that this factor is negligible. Certainly the general trend is to higher permeabilities as ID/OD is increased regardless of the variations from core to core. Also, it would seem that the authors have neglected the possibility of winding two stacked cores with common control windings in their development of optimum core shapes.

R. M. Hubbard and **M. M. Bishop**: Mr. Korn's first point is well taken. Originally it was felt that too little was known of the effect of ID/OD ratio, especially of the manner in which it varied with core material, to permit its inclusion in a generalized investigation. However, an effort has subsequently been made to add this factor to the study.

The main source of information on the manner in which the ID/OD ratio affects the dynamic performance of the core is contained in a paper by Roberts and Van Nice[1] and in a paper by Brownell.[2] These papers provide information which is applicable to one class of magnetic material, namely, grain-oriented 50% NiFe. In order to illustrate the manner in which this factor would modify the results of this particular investigation, it is necessary to introduce the concept of threshold magnetizing force $(\bar{H}_t)$ which was originally neglected in defining $\hat{R}$ as the slope of the line from the origin to the knee of the magnetization curve. This is due to the fact that, although the average magnetizing force corresponding to full reset $(\bar{H}_0)$ is substantially invariant with ID/OD, the average threshold magnetizing force $\bar{H}_t$ varies considerably (for 50% Ni-Fe). The introduction of $\bar{H}_t$ invalidates the use of the a-c magnetization curve to define $\hat{R}$ and requires the use of the type of curve obtained on some sort of adjustable reset core tester such as that described by Roberts.[3] The net result is that, in equations 59 through 64, the quantity $(\bar{H}_0-\bar{H}_t)$ must be substituted for $\bar{H}_0$ alone. Since changes in $(\bar{H}_0-\bar{H}_t)$ are inversely proportional to changes in differential permeability (as ID/OD ratio is changed), it can be seen that the constants K_3 and K_6 become variables with ID/OD. As a result, the functional relationships between ID/OD ratio and both normalized power gain and figure of merit are changed. These modified relationships, illustrated by Figs. 11 and 12, have been developed by making use of the functional relationship between differential permeability and ID/OD ratio (see Fig. 13) which is suggested in reference 1. The changes introduced by Figs. 11 and 12 are clarified by taking note of the original figures, Figs. 7(A) and (B) of the text. The new optimum functional relationship between h/b and ID/OD dictated by a compromise between load volt-amperes and figure of merit is depicted

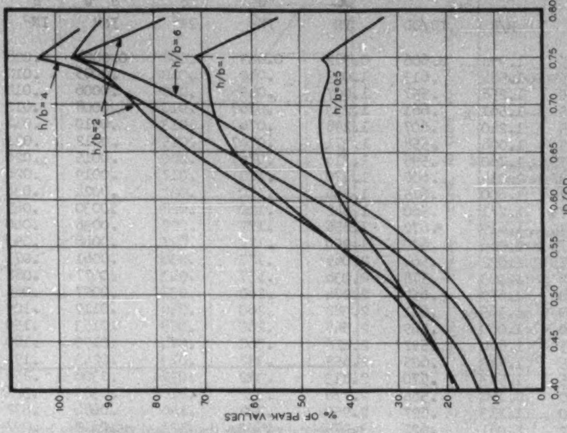
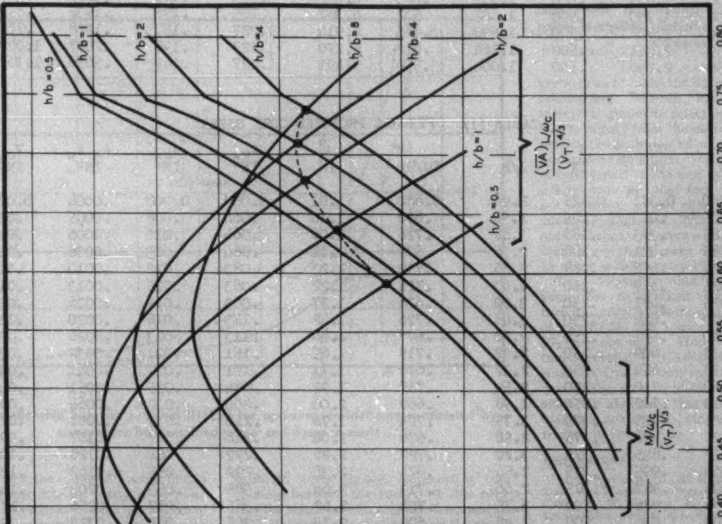


Fig. 11 (left). Load power and figure of merit (normalized) versus toroidal-core dimensional ratios h/b and ID/OD

Fig. 12 (above). Steady-state power gain (normalized) versus toroidal-core dimensional ratios h/b and ID/OD



in Fig. 14. The optimum ranges of the variables have been selected by taking note of the shape of the power-gain curve of Fig. 12.

It should be noted that the new optimum relationship between *h/b* and ID/OD (Fig. 14) is not radically different from the original one. The major difference is in the optimum ranges of ID/OD and *h/b*. Thus, the optimum range for *h/b* is shifted from the original range between one and two to the new range between two and four. One might conclude that optimum cores could logically fall anywhere between the two curves and that added emphasis on that variable involving change in permeability with ID/OD merely serves to shift the optimum ranges to higher *h/b* and higher

TABLE II OPTIONAL PROPOSED CORE SIZES

Core No.	ID IN	OD IN	H IN	H/B	ID/OD	ℓ IN	A_w IN^2	A_c IN^2	$A_c A_w$ IN^4	V_c IN^3	V_w/V_c
1	0.250	.375	0.125	1.994	0.666	0.982	0.049	0.008	0.0004	0.008	3.874
2	.250	.408	.125	1.584	.613	1.033	.049	.010	.0005	.010	2.972
3	.250	.449	.125	1.258	.557	1.097	.049	.012	.0006	.014	2.267
4	.3125	.473	.125	1.561	.661	1.233	.0767	.010	.0008	.012	4.338
5	.3125	.514	.125	1.240	.607	1.298	.0767	.013	.0010	.016	3.330
6	.3125	.560	.125	1.008	.558	1.371	.0767	.015	.0012	.021	2.614
7	.3125	.525	.1875	1.760	.595	1.316	.0769	.020	.0015	.026	2.448
8	.375	.561	.1875	2.014	.668	1.470	.110	.017	.0019	.026	3.919
9	.375	.609	.1875	1.600	.615	1.546	.110	.022	.0024	.034	3.007
10	.375	.670	.1875	1.271	.560	1.642	.110	.028	.0030	.045	2.294
11	.475	.706	.1875	1.619	.672	1.856	.177	.022	.0038	.040	4.628
12	.475	.766	.1875	1.286	.620	1.950	.177	.027	.0048	.053	3.559
13	.475	.842	.1875	1.022	.564	2.069	.177	.034	.0061	.071	2.721
14	.475	.522	.250	1.443	.578	2.036	.177	.043	.0077	.088	2.434
15	.575	.873	.250	1.679	.659	2.274	.260	.037	.0097	.085	4.066
16	.575	.941	.250	1.365	.611	2.382	.260	.046	.0119	.109	3.205
17	.575	1.047	.250	1.060	.549	2.548	.260	.059	.0153	.150	2.380
18	.575	.971	.375	1.894	.592	2.428	.260	.074	.0193	.180	2.311
19	.650	1.040	.375	1.922	.625	2.655	.332	.073	.0243	.194	2.903
20	.650	1.141	.375	1.527	.570	2.813	.332	.092	.0306	.259	2.215
21	.650	1.114	.500	2.156	.584	2.770	.332	.116	.0385	.321	2.036
22	.650	1.234	.500	1.713	.527	2.959	.332	.146	.0484	.432	1.542
23	.800	1.285	.500	2.061	.622	3.275	.503	.121	.0610	.397	2.745
24	.800	1.411	.500	1.637	.567	3.473	.503	.153	.0767	.530	2.092
25	1.000	1.492	.500	2.031	.670	3.915	.785	.123	.0966	.482	3.958
26	1.000	1.605	.500	1.652	.623	4.092	.785	.151	.1189	.619	3.121
27	1.000	1.780	.500	1.282	.562	4.367	.785	.195	.1532	.852	2.318
28	1.250	1.878	.500	1.590	.665	4.914	1.227	.157	.1928	.772	4.432
29	1.250	2.041	.500	1.264	.612	5.170	1.227	.198	.2427	1.022	3.405
30	1.250	2.246	.500	1.004	.556	5.491	1.227	.249	.3056	1.367	2.601

TABLE III OPTIONAL PROPOSED CORE SIZES

Core No.	ID IN	OD IN	H IN	H/B	ID/OD	ℓ IN	A_w IN^2	A_c IN^2	$A_c A_w$ IN^4	V_c IN^3	V_w/V_c
1	0.260	0.382	0.125	2.04	0.680	1.01	0.053	0.008	.0004	0.008	4.26
2	.285	.413	.125	1.94	.690	1.10	.064	.008	.0005	.009	4.74
3	.285	.393	.188	3.48	.726	1.06	.064	.010	.0006	.011	4.58
4	.285	.421	.188	2.76	.678	1.11	.064	.013	.0008	.014	3.53
5	.325	.456	.188	2.86	.712	1.23	.083	.012	.0010	.015	4.56
6	.325	.449	.250	4.03	.724	1.22	.083	.015	.0013	.019	4.20
7	.325	.481	.250	3.20	.676	1.27	.083	.019	.0016	.025	3.23
8	.380	.524	.250	3.48	.726	1.42	.113	.018	.0020	.026	4.58
9	.380	.561	.250	2.76	.677	1.48	.113	.023	.0026	.033	3.53
10	.438	.609	.250	2.92	.719	1.65	.151	.021	.0032	.035	4.75
11	.438	.654	.250	2.32	.670	1.71	.151	.027	.0041	.046	3.66
12	.510	.710	.250	2.50	.718	1.92	.204	.025	.0051	.048	5.16
13	.510	.762	.250	1.98	.669	2.00	.204	.032	.0064	.063	3.98
14	.525	.725	.375	3.75	.724	1.96	.216	.037	.0081	.074	4.36
15	.525	.776	.375	2.98	.676	2.04	.216	.047	.0102	.096	3.36
16	.610	.845	.375	3.20	.722	2.28	.292	.044	.0128	.100	4.65
17	.610	.905	.375	2.54	.674	2.38	.292	.055	.0162	.132	3.59
18	.700	.982	.375	2.66	.713	2.64	.385	.053	.0204	.140	4.76
19	.700	.966	.500	3.75	.724	2.62	.385	.067	.0256	.174	4.36
20	.700	1.036	.500	2.98	.676	2.73	.385	.084	.0323	.229	3.36
21	.810	1.125	.500	3.17	.720	3.04	.515	.079	.0406	.240	5.48
22	.810	1.207	.500	2.52	.671	3.17	.515	.099	.0512	.315	3.53
23	.950	1.313	.500	2.75	.723	3.56	.709	.091	.0644	.323	5.10
24	.950	1.408	.500	2.18	.675	3.70	.709	.114	.0811	.424	3.93
25	.965	1.337	.750	4.03	.722	3.62	.731	.140	.1021	.505	4.12
26	.965	1.434	.750	3.20	.673	3.77	.731	.176	.1286	.662	3.18
27	1.125	1.559	.750	3.46	.722	4.22	.994	.163	.1618	.686	4.45
28	1.125	1.672	.750	2.74	.673	4.39	.994	.205	.2037	.900	3.43
29	1.315	1.818	.750	2.98	.723	4.92	1.358	.189	.2564	.929	4.87
30	1.315	1.949	.750	2.37	.675	5.13	1.358	.238	.3228	1.218	3.76

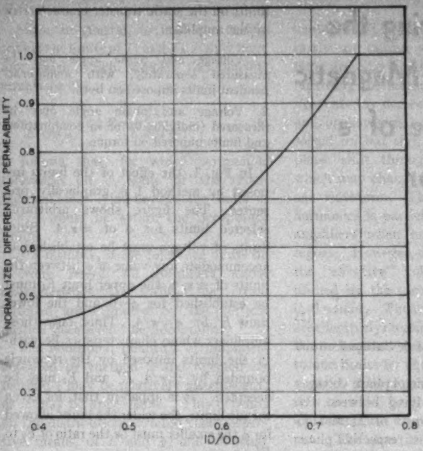


Fig.13. Differential permeability (normalized) versus the toroidal-core dimensional ratio ID/OD

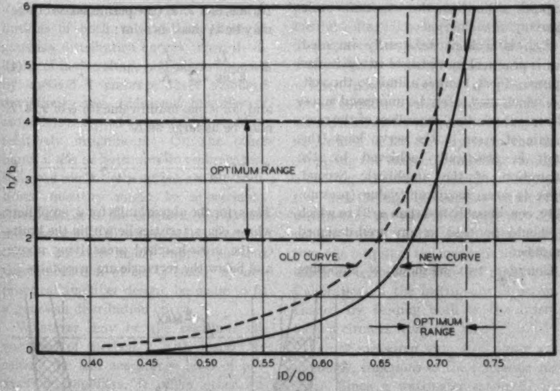


Fig. 14. Relationship between h/b and ID/OD for optimum toroidal core as dictated by a compromise between load power and figure of merit

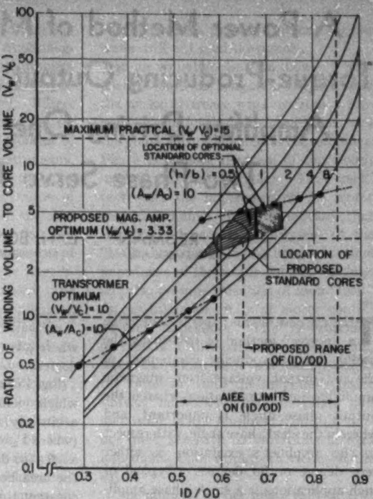


Fig. 15. Winding-volume to core-volume ratio versus the toroidal-core dimensional ratios, illustrating location of proposed standard cores

ID/OD. Thus, the important result of this extension of the original study is to point out that, if one wishes to take advantage of the increased permeability associated with the higher ID/OD ratios, he should also utilize the higher h/b ratios dictated by the optimum functional relationships of Fig. 14. Clearly, in view of the many unresolved questions associated with the relationship between the differential permeability of various materials and ID/OD ratio, a reasonable proposal for optimum standard cores should represent a compromise between the results of this study and the results of the original investigation.

To illustrate the application of the results presented in Fig. 14, two lists of possible optimum-shaped cores have been compiled. The first list (Table II) is based on the original optimum range of h/b between one and two but does not represent a rigid adherence to the original curve of Fig. 14. By adopting the practical point of view that the assumptions inherent in this analysis are such as to make the optimum functional relationship between h/b and ID/OD merely a guide in selecting cores, one is free to depart from the optimum such as to permit the inclusion of more standard ID in the core tabulation. Table III is a listing of cores based on the new optimum range of h/b between two and four and it too does not represent a rigid adherence to the new curve of Fig. 14. The location of the optional standard cores with reference to the original proposed cores is clarified by Fig. 15 which is taken from Fig. 9.

The observation that this study is not applicable to stacked cores is partially true in that the results pertaining to figure of merit (M) and steady-state power gain (G_p) are not valid. However the information associated with load power rating $(\bar{V}\bar{A})_L$ is valid. One can only say that this analysis covers the most typical core applications since it is also clear that saturable reactors, magnetic modulators, etc., are not explicitly treated in this paper.

REFERENCES

1. INFLUENCE OF ID-OD RATIO ON STATIC AND DYNAMIC MAGNETIC PROPERTIES OF TOROIDAL CORES, R. W. Roberts, R. I. Van Nice. *AIEE Transactions*, vol. 74, pt I, Nov. 1955, pp. 599-607.

2. CORE AND LAMINATION SHAPE FACTORS IN MAGNETIC AMPLIFIER DESIGN, H. R. Brownell. "Proceedings of the Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T86*, July 1956.

3. MAGNETIC CHARACTERISTICS PERTINENT TO THE OPERATION OF CORES IN SELF-SATURATING MAGNETIC AMPLIFIERS, R. W. Roberts. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

A Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor

S. J. REISMAN
NONMEMBER AIEE

E. H. BOWMAN
NONMEMBER AIEE

THE purpose of this paper is to describe a "power" method of measuring the useful, torque-producing component of the a-c output voltage from magnetic amplifiers used in applications wherein the output phase angle is important, and wherein the ideal phase angle, with respect to the amplifier's excitation, is either zero degree or 90 degrees. Typical of such applications is a single-phase amplifier used to drive one phase of a 2-phase servo motor.

In this specific application, when only single-phase power is available, the phase to one or both windings of the motor must be shifted to arrive at quadrature voltages across the windings. Usually, the phase shift is applied to one circuit only and, depending on where the phase shift is accomplished, the ideal phase angle at the output of the amplifier, with respect to the line voltage, is either zero degree or 90 degrees.

Assume that the voltage across the reference winding is held constant, that the line frequency is held constant, that the line voltage has a negligible harmonic content, and that voltages are held to those values producing motor torque within its linear range. Then the expression for the stall torque of the motor may be written

$$T = KE \sin \theta \qquad (1)$$

where T is the stall torque, K is a constant of proportionality, E is the fundamental component of voltage across the motor control winding and θ is the phase angle between the voltages on the two windings. Equation 1 may also be written

$$T = KE \sin (\pi/2 + \phi) \qquad (2)$$

where ϕ is the angle of deviation from 90 degrees.

Fig. 1 is a plot of control phase voltages which yield torques lying between the arbitrarily selected limits of T_1 and T_2 (where $T_2 = 2T_1$) against respective phase angles (as deviations from $\pi/2$). Within the area bounded by the two curves of constant torque and E_{max}, there is also a 2 to 1 ratio of permissible voltage limits for every value of ϕ, except where E_{max} imposes a limit on the maximum voltage.

E_{max} is a limit arbitrarily imposed; but is practical by virtue of two considerations. First, there is a limit to the voltage which may safely be impressed across the motor winding, regardless of the magnitude of error in the servo loop (this limit is practically achieved by the saturation of the amplifier). Second, there is some maximum value (presumably, considerably less than $\pi/2$) to which ϕ should be held in any well-designed amplifier.

Consider two methods of imposing limits on the static transfer characteristic of the amplifier:

1. Voltage and phase angle may be measured separately, with noninterdependent limits imposed on both.
2. Voltage and phase angle may be measured (individually or in combination) and limits imposed on torque.

In Fig. 1, the effect of the limits imposed by method 1 is graphically presented. The figure shows arbitrarily selected limits for ϕ of $\pm\pi/4$. Since limits of voltage must be established to accommodate any value of ϕ between the limits of $\pm\pi/4$, the upper limit E_2 must be established for $\phi = 0$ and the lower limit E' for $\phi = \pi/4$. Thus, only those amplifiers whose characteristics lie within the limits imposed by the rectangle bounded by $\pm\pi/4$, E' and E_2 may be accepted. It is apparent that, for given torque limits, the wider the limits allowed for ϕ the smaller must be the ratio of E_2 to E', or

$$\frac{T_2}{T_1} = \frac{E_2 \sin \pi/2}{E' \cos \phi_{max}} \text{ and } \frac{E_2}{E'} = \frac{T_2 \cos \phi_{max}}{T_1} \qquad (3)$$

By method 2, E (the permissible voltage) may be as small as E_1 or:

$$E = \frac{T_1}{K \sin \pi/2} = \frac{T_1}{K} \qquad (4)$$

and, for a maximum value for ϕ of $\pi/4$, E may be as large as E''.

$$E = E'' = \frac{T_2}{K \sin (\pi/2 \pm \pi/4)} = \frac{T_2}{K \cos \pi/4} \qquad (5)$$

Thus, for the above limits for ϕ, amplifiers whose characteristics lie within the limits of the cross-hatched areas lying above and below the rectangle are acceptable.

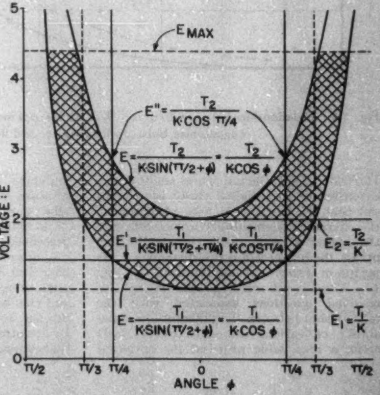


Fig. 1. Loci of $KE[\sin(\pi/2+\phi)]$ for two values of constant torque ($T_2 = 2T_1$)

Paper 58-897, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 17, 1958.

S. J. REISMAN is with the Westinghouse Electric Corporation, East Pittsburgh, Pa., and E. H. BOWMAN is with the Sandia Corporation, Albuquerque, N. Mex.

Since, in general, to hold the torque within the limits of T_1 and T_2, E may have any value which satisfies the limiting equations of

$$E=\frac{T_1}{K\cos\phi} \text{ and } E=\frac{T_2}{K\cos\phi} \qquad (6)$$

it follows that, for wider permissible tolerances on ϕ, some of the cross-hatched areas lying beyond the limits of $\pm\pi/4$ would be within the acceptance limits.

For instance, if the tolerance limits on ϕ were $\pm\pi/3$, all amplifiers whose characteristics lie within the cross-hatched areas limited by the vertical dashed lines would be acceptable.

The additional quantity of amplifiers accepted by method 2 and rejected by method 1 may be relatively insignificant; or it may be appreciable. This depends on the distributions, around their respective means, of E and ϕ, and on what percentage of the total quantity of production amplifiers has characteristics which lie within the limits of E and ϕ required, by method 1, to satisfy the limits placed on torque.

For instance, assume that the distributions of both voltage and phase fit gaussian distribution curves; then if the limits of both voltage and phase imposed by method 1 embrace three standard deviations, any additional quantity of amplifiers accepted by method 2 would be relatively insignificant. On the other hand, if one or both were to embrace two standard deviations or less, such additional quantity might be appreciable. Obviously, it is the function (cos ϕ) and not the angle (ϕ) which is of direct interest; but it is highly improbable that the distribution of cos ϕ could, by any practical amplifier design, be made to fit a gaussian distribution curve.

Whatever may be the resultant advantages of method 2 with respect to raising the test acceptance level of production amplifiers, it offers other advantages. The high degree of distortion in the a-c output voltage of most magnetic amplifiers poses a difficult problem regarding the measurement of such voltage with any degree of accuracy. Filters and/or harmonic analyzers may be resorted to, but this poses the problem of phase shift through such instruments, which may change the shape of the voltage waveform. The nuisance value of harmonics is particularly great in servo amplifiers when measuring in the null region. However, direct measurement of the "effective" voltage can be accomplished by the use of a wattmeter type instrument. Thus, the voltage and phase are effectively measured "in combination", limits imposed are directly related to torque limits by equations 1 and 2, and method 2 is applicable. Additionally, it is possible, by method 2, to obtain a more accurate measurement of null output. The practical limitation of mismatch (in cores, turns, resistors, and rectifiers) dictates that the combined harmonic output voltage of push-pull magnetic amplifiers at the null of the fundamental is never zero. However, motor torque passes through zero with the fundamental voltage if no harmonics are present in the reference winding, and this is indicated by the wattmeter.

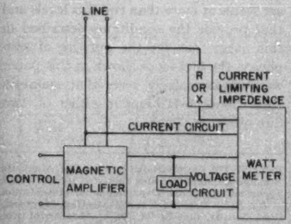


Fig. 2. Block diagram of a power-measuring circuit for a magnetic amplifier

The "power" method of measuring torque employs a reference current, representative of motor reference winding voltage, which is fed directly from the line supplying the magnetic amplifier through a limiting impedance (resistive or capacitive) and through the current circuit of a wattmeter, as shown in Fig. 2. The voltage circuit of the instrument is fed from the output of the magnetic amplifier, which is terminated into its rated load. Calibration of the instrument is accomplished by feeding both of the instrument's circuits from the line.

Actual measured values of power are, of course, functions of the reference current. Since a wattmeter reads EI cos ϕ (where ϕ is the phase angle between the voltage and current), it is necessary that the type of current-limiting impedance be chosen to match the nominal phase angle of the amplifier's output voltage with respect to the line. Hence, if the amplifier's output is nominally in quadrature with the line the reference current is fed through a capacitor; and if the amplifier's output is nominally in phase with the line the reference current is fed through a resistor. When measuring the phase-reversible output of a push-pull amplifier, provisions must be made for reversing the connections of either the current or voltage circuit of the meter as the amplifier's output reverses phase, in order that the meter will not read downscale.

In an actual application, power is supplied to both the magnetic amplifier and the motor reference winding from a 115-volt 400-cps (cycles per second) line, and the phase shift is effected in the amplifier. Fig. 3 is a block diagram of the circuit used to test this amplifier. The "load" represents the impedance of the motor control winding. It is significant that this method permits adjustment of the "load" impedance to the actual nominal impedance of the motor control winding, and also permits variation of such "load" to the tolerance limits of motor control winding impedance. This is especially pertinent in this particular application, where the phase shift is accomplished in the amplifier, since the amplifier's phase shifting network makes the phase and amplitude of the amplifier's output voltage more sensitive to variations in load impedance than would otherwise be the case.

The nominal carrier phase shift of ±90 degrees through the amplifier necessitates the use of a capacitor, rather than a resistor, to limit the wattmeter's reference current. The value of the capacitor is set at 0.2075 μf (microfarad), limiting the current (at 115 volts 400 cps) to 60 ma (milliamperes), which is the rating of the wattmeter's current circuit. To calibrate the meter, a resistance of 1,917 ohms is substituted for the capacitor, and both the voltage and current circuits of the meter are connected to the line. Thus, reference "power" (which is reference torque times a constant of proportionality) is

$$W=EI\cos\phi=(115 \text{ volts})(0.060 \text{ ampere})(1)=6.9 \text{ watts} \qquad (7)$$

For calibration at other points, the reference current should be held constant (to duplicate the conditions for use of the

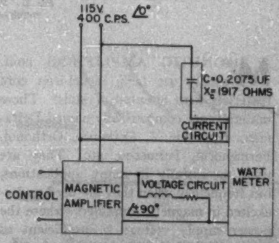


Fig. 3. Block diagram of a power-measuring circuit for a specific magnetic amplifier with quadrature voltage at the amplifier output

instrument) and the potential across the wattmeter's voltage circuit varied. Thus, the wattmeter reading becomes a linear function of the varying voltage. A check for zero power indication should be made by substituting the capacitor for the resistor.

Then, with the meter's current circuit connected through the capacitor, the amplifier terminated into its "load," the voltage circuit of the wattmeter connected across the "load," and reference current (at 115 volts 400 cps) fixed at 60 ma, the wattmeter's "power" reading will be proportional to the fundamental output voltage of the amplifier times the cosine of the phase angle (nominally zero) between this voltage and the reference current. The torque, then, is the wattmeter reading times a constant of proportionality.

$$T=KW=KE\cos\phi \qquad (8)$$

The fundamental output voltage of the magnetic amplifier is the wattmeter reading divided by the product of the reference current and the cosine of the phase angle

$$E=\frac{W}{0.06\cos\phi} \qquad (9)$$

and the "effective voltage" (which is $E\cos\phi$) is the wattmeter reading divided by the reference current

$$E_{eff}=\frac{W}{0.06} \qquad (10)$$

Thus, the true value of torque-producing voltage can be calculated, see equation 10, and this times a constant of proportionality yields the value of torque, as determined by the carrier phase shift through the amplifier and the amplifier's fundamental output voltage, which in turn depends upon the amplifier's control input and voltage gain.

The input impedance of the wattmeter's voltage circuit must be high enough to avoid a loading effect on the amplifier; the current sensitivity must be high enough to avoid loading the line beyond its capabilities; and, in the case of a capacitor as the limiting impedance, the resistance of the current circuit must be low enough, relative to the capacitive reactance, that the power factor of the circuit is not appreciable. A power factor of 0.01 at 400 cps is practicable. In the case of resistance current-limiting, a minimum power factor of 0.99 at 400 cps should be attainable without difficulty. The instrument used in the above application has a current circuit resistance at the 60-ma level of 0.75 ohm, and the input impedance of the voltage circuit is 1 megohm of resistance shunted by 25 micromicrofarads.

If it is desired to measure the transfer characteristic of the amplifier under variations in line voltage and/or frequency, the reference current must be adjusted for each condition, or limits of the transfer characteristic must be adjusted to accommodate for variation of the reference current with variations in line voltage and (in the case of capacitor current-limiting) frequency.

There are, of course, electromechanical methods of measuring the torque-producing output voltage of an a-c magnetic servo amplifier and, in applications where the phase angle of the output voltage is not pertinent, they appear to have considerable merit. Such methods usually involve the use of a motor as the amplifier load, working it against a counter-torque obtained from springs, weights or another motor. Problems common to these methods are motor shaft coupling and alignment, and the variation in motor torque with respect to the motor's electric poles and shaft position in a multipole motor. Additional problems encountered with the use of springs or weights are calibration and reading accuracy. These latter problems should not be serious when the counter-torque is obtained from another motor.

In applications wherein the phase angle of the output voltage is pertinent, however, the wattmeter method appears to offer definite advantages, especially with regard to the use of a dummy load which, as pointed out earlier in this paper, offers the flexibility of adjusting the load impedance to any desired value. This flexibility is of special value in applications such as the particular one discussed herein, due to the sensitivity of the magnitude and phase angle of the amplifier's output voltage to variations in load impedance.

Reactions of Bistable B-H Loops on Magnetic Amplifiers

R. E. MORGAN
MEMBER AIEE

MAGNETIC AMPLIFIERS commonly use 50% nickel-iron core materials with oriented crystals. These materials are commercially named Deltamax, Hypernic V, Orthonik, Orthonol, Permenorm, Permeron, etc. They are used successfully in many applications, but frequently when they are used underexcited in magnetic amplifiers (where the power supply voltage is insufficient to drive the core from negative to positive saturation), instability and slow response time occur.

The instability of magnetic amplifiers caused by 50% oriented nickel-iron cores has previously been described,[1] and the characteristics of these cores have been explained.[2-4] This paper presents measured **B-H** loops (kilogausses versus oersteds) with some characteristics of bistable minor **B-H** loops, and shows that instability (Fig. 1) and slow response time of magnetic amplifiers can be caused by the bistable **B-H** loops (Fig. 2). The use of a pulse to preset the core of the magnetic amplifier is described. The preset pulse forces the core to operate on a single minor **B-H** loop and causes the magnetic amplifier to have a stable characteristic (Fig. 3) and fast response (Fig. 4).

The term "bistable **B-H** loop" is used to indicate that a minor **B-H** loop of the core may remain stable at two flux position levels for the same external conditions (voltage and ampere-turns). Actually some cores have minor loops that are stable at more than two flux levels and that provide the conditions described in this paper, but for the purpose of simplicity, the **B-H** loops shown in this paper were obtained from cores that maintain these minor **B-H** loops in either of two flux position levels.

Paper 57-776, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 9, 1957; made available for printing May 5, 1958.

R. E. MORGAN is with the General Electric Company, Schenectady, N. Y.

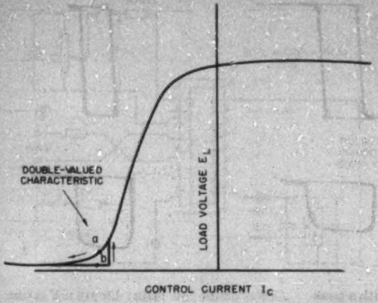


Fig. 1 (left). Transfer characteristic of a magnetic amplifier with an instability

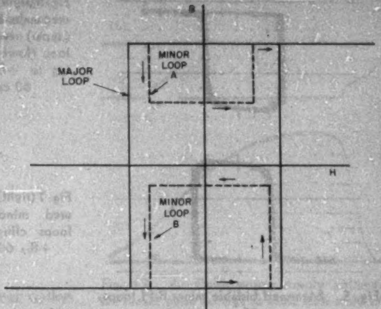


Fig. 2 (right). Idealized B-H loops

Nomenclature

e_A = primary winding A voltage, volts, instantaneous
e_B = primary winding B voltage, volts, instantaneous
E_C = control voltage, volts, average
e_L = load voltage, volts, instantaneous
E_L = load voltage, volts, average
$E_{L,\max}$ = maximum load, volts, average
$E_{L,\min}$ = minimum load, volts, average
$e_{P,P}$ = preset pulse voltage, volts, instantaneous
e_S = supply voltage, volts, instantaneous
I_B = bias current, amperes, average
i_C = control current, amperes, instantaneous
I_C = control current, amperes, average
i_L = load current, amperes, instantaneous
I_L = load current, amperes, average
$i_{P,P}$ = preset pulse current, amperes, instantaneous
rect_1 = "half wave" circuit gate rectifier
rect_2 = saturable transformer load rectifier
S_1 = control circuit switch
S_2 = preset pulse circuit switch

Bistable B-H Loops

Bistable minor **B-H** loops are shown in Figs. 5 through 10 for a 50% nickel-iron core. These **B-H** loops show some characteristics that exist in bistable loops. Taken from a single core, they are not to be interpreted as typical or average. From 12 cores selected at random, 8 of the cores have **B-H** loops similar to those of Figs. 5 through 10, while the remaining 4 cores have symmetrical minor **B-H** loops (minor loops showed no signs of saturation). The a-c **B-H** loops (Figs. 5 through 8) were obtained by using a sine wave flux at 60 cps (cycles per second).

Bistable a-c minor **B-H** loops are shown in Fig. 5. The upper portion of the upper loop remains at the positive residual flux level ($+B_r$), and the lower portion of the lower loop remains at the negative residual flux level ($-B_r$). The upper loop occurs by applying $+25$ oersteds of direct current to the reactor, removing the direct current, and applying a sine wave flux at 60 cps to the reactor. The lower loop occurs in the same manner except that -25 oersteds are used. Figs. 6 through 8 show the a-c major **B-H** loop and several amplitudes of a-c minor **B-H** loops that cling to the positive residual flux density level $+B_r$.

Bistable d-c minor **B-H** loops are shown in Fig. 9. **B-H** loop A (dashes) and **B-H** loop B (dots) were obtained by cycling, very slowly, between $+0.13$ and -0.13 oersteds for many cycles. Loop A was started at $+B_r$ and loop B was started at $-B_r$. The bistable loops A and B are shown with the d-c major **B-H** loop (solid line) and two other minor loops that cling to $+B_r$.

The **B-H** characteristics (d-c) of Fig. 10 illustrate the "memory" nature of the core. The solid curve is the major d-c loop, the dash curve is a **B-H** characteristic originating from positive saturation, and the dotted curve originates from negative saturation. During the time that flux is moving away from the original saturation flux level, the characteristic follows close to the major **B-H** loop although the magnetomotive force H was removed ($H=0$ or even slightly reversed) and H reinstated. During the time that flux is moving toward the original saturation level, the **B-H** characteristic is farther inside the major d-c loop (dashed curve with H positive) than during the time that flux is moving away from the original saturation level. Fig. 10 shows that the core remembers the previous saturation polarity with magnetomotive force H at $H=0$ for various levels of flux density B.

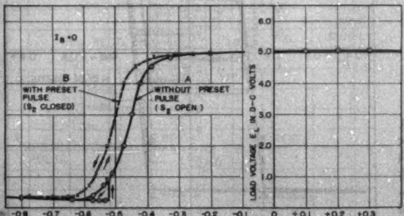


Fig. 3. Transfer characteristic of a magnetic amplifier, circuit in Fig. 11, with and without preset pulse

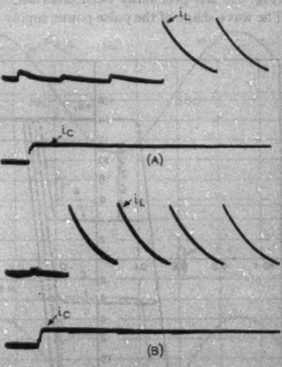


Fig. 4. Response time of magnetic amplifier of Fig. 11. A—Load voltage e_L and control voltage e_C without preset pulse. B—Same as (A) but with preset pulse

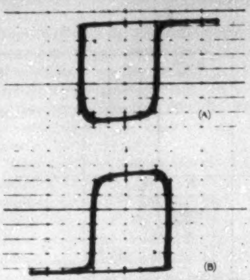


Fig. 5. Measured bistable minor B-H loops, 60 cps

Reaction of the Magnetic Amplifier

The bistable **B-H** loops react upon the magnetic amplifier by causing a double valued portion of the transfer characteristic (curve *A*, Fig. 3) and causing the output to respond slowly to the application of control current; see Fig. 4(C). These undesirable characteristics are corrected by a preset pulse as shown in curve *B* of Figs. 3 and 4(B). The doubled-value characteristic and the slow response can occur in any of the magnetic amplifier circuits[5] ("bridge," "center-tap," etc.), but the "half-wave" magnetic-amplifier circuit of Fig. 11 was chosen to present these characteristics because of the circuit's simplicity. The "half-wave" magnetic amplifier uses a pulse power supply instead of the usual sine wave supply.[5] The use of a pulse power supply for this circuit (Fig. 11) has previously been described.[6] The wave shape of the pulse power supply voltage is shown in Fig. 12(A) with a peak voltage $e_S = 100$ volts and a pulse width of 100 microseconds (approximately 2 degrees of a 60-cps wave). The preset pulse, Fig. 12(B), is compared to the supply pulse, Fig. 12(A), and the supply pulse voltage e_S is shown, Fig. 12(C), expanded in the time scale.

Fig. 6 (right). Measured major B-H loop (upper) and minor loop (lower) clinging to $+B_r$, both 60 cps

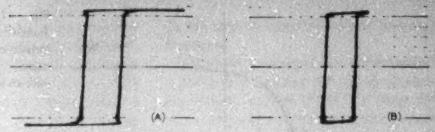
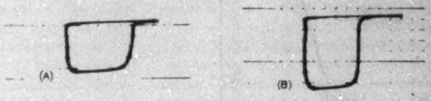


Fig. 7 (right). Measured minor B-H loops clinging to $+B_r$, 60 cps



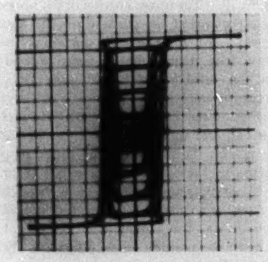


Fig. 8. Measured major and several minor B-H loops superimposed, 60 cps

CORRECTION OF THE DOUBLE-VALUED CONTROL CHARACTERISTIC

Curve *A* of Fig. 3 is the characteristic of the magnetic amplifier of Fig. 11 without bias ($I_B=0$) and without the preset pulse (S_2 open). The double-valued control characteristic (at $E_L<1.0$ volt) is caused by the bistable **B-H** loops as previously explained.[1] The application of the preset pulse (S_2 closed, Fig. 11) removes the double-valued characteristic as shown by curve *B* of Fig. 3. The preset pulse is applied (for 500 microseconds) shortly after the power pulse (Fig. 12) and drives the core (Fig. 11) positive to +30 oersteds, and transfers the core to the upper bistable **B-H** loop (Figs. 2 and 5). The preset pulse is applied for a longer interval and it provides more volt-seconds than the main power pulse in order to insure preset for all normal conditions of control, and the resistor (2,000 ohms, Fig. 11) limits the peak current of the preset pulse. The preset pulse forces the core into positive saturation after each power pulse. A winding, separate from the gate winding (power winding), is used for

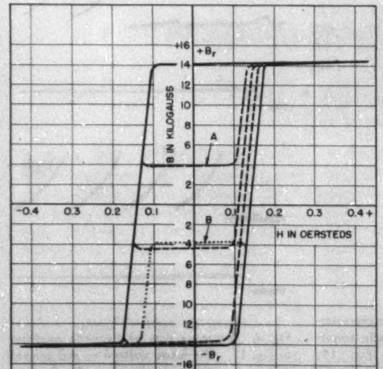


Fig. 9 (left). Measured major and minor B-H loops, d-c, including bistable minor loops A and B

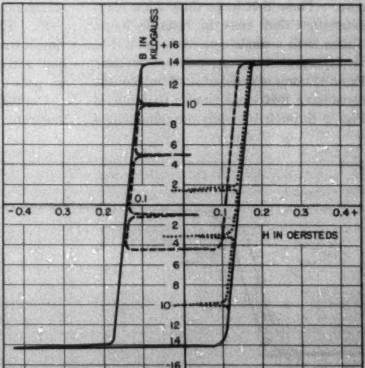


Fig. 10 (right). Measured major and minor B-H characteristics, d-c, showing memory property of the core

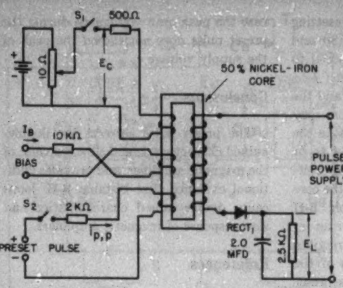


Fig. 11. Half-wave magnetic-amplifier circuit used to investigate the reaction of the bistable **B-H** loops

preset so that the preset current $i_{p,p}$ does not flow in the load circuit. While the core is saturated (+30 oersteds) by the preset pulse, the supply voltage $e_S=0$, Fig. 12(A) and the reaction of the preset pulse does not shift the characteristic of the magnetic amplifier very much nor does it change the gain.

The characteristic of the circuit and core without pulse preset and with a sine wave power supply (not shown) is approximately the same as that with the pulse power supply. The pulse preset principle can be used with a sine wave power supply (shown later) but the phasing of the preset pulse presents a problem. If the preset pulse lags the end of the conducting half-cycle of the power supply, there is insufficient time for resetting the core in most circuit applications and the minimum load voltage E_L is excessive. The pulse power supply provides ample time for reset between the end of the preset pulse and the following power pulse; see Fig. 12.

Correction of Slow Response

The reaction of the bistable **B-H** loops upon the magnetic amplifier causes slow response to control signals that drives the load voltage E_L from its minimum value to intermediate values (such as $E_L=3.0$ volts in Fig. 3). Response time up to 2 seconds has been observed with the circuit of Fig. 11 with a marginal control voltage of $E_C=-0.52$ volt. This condition being unusual and extreme is not shown in this paper. The response time shown in Fig. 4(A) is typical for the circuit of Fig. 11 with bias current $I_B=2.0$ milliamperes (equivalent of $E_C=-1.0$ volt in Fig. 3), switch S_2 open, and the application of control voltage $E_C=+0.55$ volt (providing net control of $E_C=-1.0+0.55=-0.45$ volt finally producing load voltage $E_L=3.0$ volts). The response time is slightly longer for the application of control voltage $E_C<0.55$ volt (producing $E_L<3.0$ volts).

The slow response time is caused by the time required to transfer from the lower bistable minor **B-H** loop (Figs. 2 and 5) to the upper minor **B-H** loop. This is corrected by the preset pulse. The preset pulse transfers the core from the lower minor to the upper minor **B-H** by the application of +30 oersteds by a method similar to that by which the minor loops of Fig. 5 are transferred. The same preset pulse that corrected the double-valued control characteristic corrects the slow response, as shown in Fig. 4(B). The preset amplitude of +30 oersteds is not critical, and the minimum amplitude depends upon the pulse length of the preset pulse and the particular core being applied.

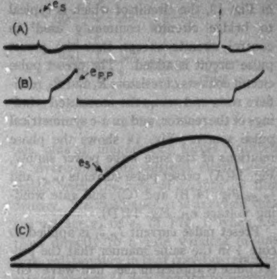


Fig. 12. A—Pulse power supply voltage. B—Preset pulse voltage to same time scale of 10-millisecond oscillogram length. C—Pulse power supply voltage with time scale expanded; 100-microsecond oscillogram length

Application of Preset Pulse to Magnetic Amplifiers Using Sine Wave Power Supply

The preset pulse, described in the foregoing, is applicable to magnetic amplifiers that use a sine wave power supply. A bridge type of magnetic amplifier is shown

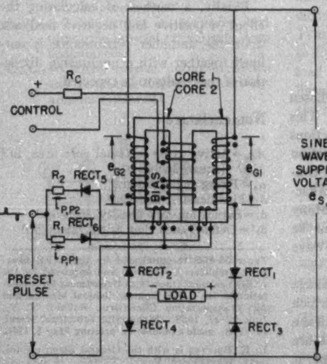


Fig. 13 (left). Full-wave bridge magnetic amplifier with preset pulse

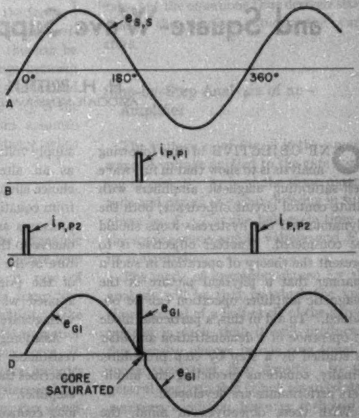


Fig. 14 (right). Phase relation and wave shapes of:

A—Supply voltage $e_{S,S}$
B—Preset pulse current $i_{P,P1}$
C—Preset pulse current $i_{P,P2}$
D—Gate voltage e_{G1} from circuit of Fig. 13

in Fig. 13, the circuit of which is typical to bridge circuits commonly used[5] in control circuits except that the preset pulse circuit is added. The preset pulse circuit consists of resistors R_1 and R_2, rectifiers rect_3 and rect_4, the associated windings of the reactor, and an a-c symmetrical pulse source. Fig. 14 shows the phase relations of the sine wave power supply, Fig. 14(A), preset pulse currents $i_{P,P1}$ and $i_{P,P2}$ Fig. 14(B) and (C), and gate winding voltage e_{G1}, Fig. 14(D).

Preset pulse current $I_{P,P1}$ is applied to core 1 in the same manner that the preset pulse is applied in the "half-wave" circuit of Fig. 11, and preset pulse current $i_{P,P2}$ is applied to core 2 likewise, except that current $i_{P,P2}$ is 180 degrees out of phase with current $i_{P,P1}$. The preset pulse is applied at the end of the conducting half-cycle of the corresponding gate circuit. Consider the reaction of the preset pulse current $i_{P,P1}$ upon core 1. Pulse current $i_{P,P1}$ is applied at 180 degrees of the supply voltage $e_{S,S}$ cycle, and pulse current $i_{P,P1}$ is applied for a sufficient time to saturate core 1 if core 1 is not already saturated at 180 degrees. Fig. 14(D) shows gate voltage e_{G1} when the circuit of Fig. 13 is producing minimum output ($e_{G1}=e_{S,S}$ during 0 degree to 180 degrees) and the core saturates during the preset pulse. In the case of a pure resistive load (Fig. 13), the core's maximum resetting volt-seconds ($\int e_{G1}\,dt$ between 180 and 360 degrees) are approximately equal to the maximum setting volt-second ($\int e_{G1}\,dt$ between zero and 180 degrees) and the preset pulse needs to force the core only the small amount necessary to make the setting volt-seconds at least equal to or slightly greater than the resetting volt-seconds. The preset pulse forces the core to operate on the positive bistable **B-H** loop. In the case of a resistive or an inductive load (commutated by rectifier rect_3 and rect_4), the volt-seconds of the gate winding ($+\int e_{G1}dt$) during the preset pulse need to be 5% or less of the volt-seconds of the gate winding between zero and 180 degrees. If the load in Fig. 13 had been a capacitive load, the possible resetting volt-seconds of the gate winding ($-\int e_{G1}dt$) between 180 and 360 degrees would be greater than the resetting volt-seconds for a resistive load and the pulse preset would have to force a greater flux change in the core for a capacitive load than a resistive load. Although in some applications, such as capacitive loads, the peak-induced gate voltage e_{G1} may be excessive during the preset pulse, most applications using resistance loads or inductive loads can obtain good results with a preset pulse length of about 5 degrees of the supply voltage $e_{S,S}$ cycle, in which case the peak gate voltage e_{G1} during the preset pulse does not exceed the peak of the supply voltage $e_{S,S}$.

Conclusions

The preset pulse corrects the double-valued characteristic and slow response of the magnetic amplifier, and provides additional evidence that bistable **B-H** loops cause double-valued characteristics and slow response of magnetic amplifiers.

References

1. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, S. B. Batdorf, W. N. Johnson. *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 223–28.

2. Flux Reversal in Magnetic Amplifier Cores, F. J. Friedlaender. *Ibid.*, vol. 75, pt. I, July 1956, pp. 268–78.

3. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, Harold W. Lord. *Ibid.*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 721–28.

4. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, H. W. Lord. *Ibid.*, pp. 85–88.

5. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

6. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 229–40.

7. The Biased Rectifier Amplifier—A Pulse Magnetic Amplifier, R. E. Morgan. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 584–90.

Characteristics of Magnetic Amplifiers Utilizing Square-Loop Core Material and Square-Wave Supply Voltage

H. H. BRITTEN
ASSOCIATE MEMBER AIEE

ONE OBJECTIVE of the following analysis is to show that in full-wave self-saturating magnetic amplifiers with finite-control circuit impedance, both the dynamic and d-c hysteresis loops should be considered. Another objective is to present the theory of operation in such a manner that a physical picture of the magnetic amplifier operation can be obtained. To aid in this, a particular mode of operation of a demonstration amplifier is studied by a step by step procedure. Finally, equations predicting the amplifiers performance are developed.

With these objectives in mind, the supply voltage to the amplifier was chosen as an alternating square wave. This choice eliminated trigonometric functions from equations and simplified integration processes and, therefore, was more conducive to the formation of a physical picture of the amplifier's operation. Many of the principles thus derived can be applied where other than square-wave voltages are used.

Assuming infinite rectifier leakage resistance, a series is developed which describes the transient performance of the amplifier. For response times that are long compared to the half-cycle time, it will be shown how a time constant is derived from this series. Equations for power gain and dynamic power gain are derived in terms of amplifier constants.

The effect of finite rectifier leakage resistance is analyzed and equations expressing its influence on amplifier performance are developed.

Finally, a method of calculating the effect of positive and negative feedback upon the amplifier performance is outlined together with a qualitative discussion of the results to be expected.

Nomenclature

A_{Fe} = Active cross-sectional core area, in.^{2} (square inches)
α_s = Firing angle, radians
β = Flux density, lines/in.^{2}
β_s = Saturation flux density, lines/in.^{2}
β_s' = Defined such that $\beta_s'<\beta_s$, lines/in.^{2}

Paper 58-856, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing May 5, 1958.

H. H. Britten is with the General Electric Company, Waynesboro, Va.

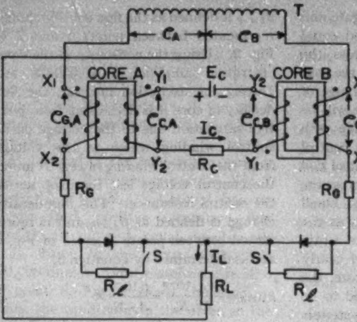


Fig. 1 (left). Center-tap magnetic amplifier circuit

$\beta_{B,e,F}$ = Change in flux density required in core B during its excitation interval before saturation occurs, lines/in.^{2}

$\beta_{A,e,R}$ = Change in flux density from positive saturation of core A during the excitation interval of core B, lines/in.^{2}

$\beta_{A,s,R}$ = Change in flux density from positive saturation of core A during the saturation interval of core B, lines/in.^{2}

β_F = Change in flux density required for saturation of a core during its excitation interval, lines/in.^{2}

β_R = Total change in flux density from positive saturation of a core during excitation and saturation interval of the other core, lines/in.^{2}

E = Magnitude of square-wave voltage, volts

e_A, e_B = Square-wave voltage applied to gate windings of core A and core B respectively, volts instantaneous

$e_{C,A}, e_{C,B}$ = Voltage of control coils of cores A and B respectively, volts instantaneous

E_c = Control voltage, volts

$e_{G,A}, e_{G,B}$ = Gate voltage of cores A and B respectively, volts instantaneous

H = Magnetic force, ampere-(amp) turns/in.

$H_{c,1}$ = Coercive force of d-c hysteresis loop, amp-turns/in.

$H_{c,2}$ = Coercive force of dynamic hysteresis loop, amp-turns/in.

$i_{L,e}, i_{L,s}$ = Load current in excitation and saturation interval respectively, amp instantaneous

$i_{c,e}, i_{c,s}$ = Control current during excitation and saturation interval respectively, amp instantaneous

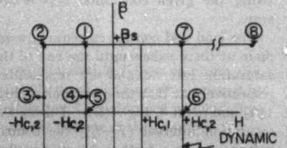


Fig. 2. Square-loop core characteristics

I_c = Control current, amp average

I_L = Load current, amp average

l_{Fe} = Mean length of magnetic circuit, in.

N_c = Control turns

N_G = Gate turns

P_G = Power gain

P_D = Dynamic power gain, sec.$^{-1}$

R_L = Load resistance, ohms

R_c = Control resistance, ohms

R_l = Rectifier leakage resistance, ohms

R_G = Gate winding and forward rectifier resistance, ohms

t_x = Excitation interval, sec.

T = Half-cycle time, sec.

$(T-t_x)$ = Saturation interval, sec.

Circuit

The magnetic amplifier circuit that will be used in this analysis is the center-tapped circuit of Fig. 1. This amplifier has two isolated cores, A and B. These cores have identical gate windings (between X_1 and X_2) and these are energized by the voltages e_A and e_B from transformer T. The resistance of the gate winding is normally R_G, however, if the forward characteristic of the rectifier is approximated by a linear resistance, this can be included in R_G. The windings between Y_1 and Y_2 on each core are interconnected to form the control winding. R_L is the load resistance and R_l is an assumed linear rectifier leakage resistance when the switches S are closed.

Assumptions

It is assumed that the dynamic hysteresis loop will exhibit square-loop characteristics for a constant rate of change of flux density. The term "dynamic hysteresis loop" is defined as the relationship between magnetic induction β and apparent magnetizing force H. The apparent magnetizing force includes that required by the d-c hysteresis loop, taken over the same operating flux densities, plus all magnetizing forces required to overcome eddy current effects associated with the rate of change of flux density. In this paper, it is assumed that the voltage e_A or e_B of Fig. 1, when applied directly to a gate winding of one core, will produce in that core the required rate of change of flux density to produce the dynamic hysteresis loop of Fig. 2. The d-c loop presumes a relatively low rate of change of flux density. In the following analysis, it will be assumed that during the portion of the cycle when the flux density of a core is changing due to the control voltage alone, the action will take place along the d-c loop. Another assumption regarding the curves of Fig. 2 is that they can only be traversed along constant H or β lines.

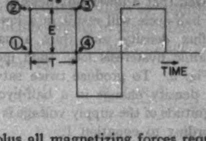


Fig. 3 (right). Square-wave voltage supply to gate of core B

There is assumed to be no leakage flux between the gate and control windings. The inductance of the gate winding during the saturation interval ($t_x<t<T$) is neglected. During the excitation interval ($0<t<t_x$) the voltage drop over the load and gate coil resistance, caused by the excitation current, is neglected.

The rectifiers are assumed to have a linear forward resistance (which is included in R_G) and an infinite leakage resistance until switches S are closed after which it is assumed to be the linear resistance R_l.

These approximations will produce second-order variations from observed data, but the equations thus derived show clearly the relations among the variables.

Step-by-Step Analysis of an Amplifier

The voltage e_B of Fig. 3 is the square-wave of voltage applied to the gate winding of core B. e_A is of the same magnitude and period as e_B but differing in time phase by T. T is the half-cycle time of the square-wave supply voltage and E is its magnitude.

The mode of operation chosen for the analysis will be that in which the volt-time integral of a half-cycle of the supply voltage is such that it will produce a flux density change in one core equal to twice saturation flux density. In the general analysis this is not assumed as a necessary condition, however, it should be realized that a lower flux density change per half-cycle will result in a lower rate of flux density and thus a narrower dynamic hysteresis loop than indicated in Fig. 2. To produce twice saturation flux density change in a half-cycle, the magnitude of the supply voltage is chosen according to equation 1.

$$E=\frac{1}{T}\int_0^T e_B dt=\frac{N_G A_{Fe}}{T}\int_{-\beta_s}^{+\beta_s} d\beta 10^{-8}$$
$$=\frac{2\beta_s A_{Fe} N_G 10^{-8}}{T} \quad (1)$$

The constants chosen for the demonstration amplifier are as follows:

$E=100$ volts
$H_{c,1}=0.25$ amp-turns/in.
$H_{c,2}=0.75$ amp-turns/in.
$\beta_s=100{,}000$ lines/in.^{2}
$A_{Fe}=0.625$ in.^{2}
$N_G=100$ turns
$N_c=100$ turns
$T=0.00125$ sec.
$l_{Fe}=8.0$ in.
$R_c=200$ ohms
$R_L=15$ ohms
$R_G=5$ ohms

The criterion that will be used to ascertain whether the amplifier is operating at a steady-state condition is as follows: If the conditions existing in both cores at the beginning of a half-cycle of supply voltage are exactly interchanged at the beginning of the next half-cycle, then the amplifier is in a steady-state operating mode. For the specific operating mode analyzed by the step-by-step method, wave shapes of load and control currents will be examined and the average values of these currents will be recorded.

The convention of polarities is as follows: When voltages are applied to a winding such that the polarity marked end is positive, the flux density change will be toward positive saturation. Similarly, when current flows into a winding at the polarity marked end, the resulting magnetic force will be positive.

On Figs. 2, 3, and 4 are marked specific points that will be referred to in following the amplifiers performance throughout a half-cycle of operation. For the specific operating mode to be examined, it will be assumed that $E_c=-8.0$ volts, and at the beginning of a positive half-cycle of voltage to the gate winding of core B (point 1, Fig. 3), conditions in core A are described by point 1, Fig. 2 ($\beta=+\beta_s$, $H=-H_{c,1}$) and core B by point 5, Fig. 2 ($\beta=0$, $H=-H_{c,1}$). As the voltage increases to point 2, Fig. 3, the current in the gate winding of core B and the load increases until the magnetic force of core B is at point 6, Fig. 2 ($\beta=0$, $H=+H_{c,2}$). At this time, the supply voltage appears across the gate winding of core B and $e_{G,B}=E$. Since the gate and control turns have been assumed equal $e_{c,B}=e_{G,B}=E$. The voltage across the control coil of core $B(e_{c,B})$ minus the control voltage E_c (in the normal range of the amplifier, the control voltage will be negative and thus will be additive with $e_{c,B}$) minus the drop across the control resistance appears across the control coil of core A. If the voltage drop across the control resistance is relatively small compared to $e_{c,B}$, the voltage across the control coil of core A will produce a rate of change of flux density in core A nearly equal to that in core B. Therefore, the flux density in core A is assumed to be changing along the dynamic hysteresis loop. Thus, the current in the control circuit increases until the magnetic force of core A is at $-H_{c,2}$ and is described by point 2, Fig. 2 ($\beta=+\beta_s$, $H=-H_{c,2}$). The control current ($i_{c,x}$) at this moment (point 1, Fig. 4(B)) is that required in the control coil to overcome the coercive force of the dynamic loop of core A. Its value is given by equation 2.

$$i_{c,x}=-\frac{l_{Fe}}{N_c}H_{c,2} \quad (2)$$

Using the constants for the demonstration amplifier, $i_{c,x}=-0.06$ amp. The load current ($i_{L,x}$) at this time (point 1, Fig. 4(A)) is that required in the gate coil of core B to overcome the coercive force of core B ($H_{c,2}$) plus the current reflected from the control coil of core B in overcoming the coercive force of core A. The excitation current at this point is given by equation 3. Again applying the given constants $i_{L,x}=0.120$ amp.

$$i_{L,x}=\frac{2H_{c,2}l_{Fe}}{N_G} \quad (3)$$

The load and control currents will remain constant at the foregoing values until core B saturates (point 7, Fig. 2). The time required for this to occur is called the excitation interval (t_x) and can be obtained by rearranging equation 1.

$$t_x=\frac{\beta_{B,x,F}A_{Fe}N_G}{E}(10^{-8}) \quad (4)$$

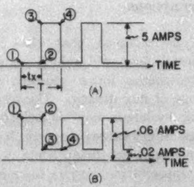


Fig. 4. Instantaneous control current

$\beta_{B,x,F}$ is defined as the flux density change represented between points 6 and 7 of Fig. 2. Using the constants of the demonstration amplifier, $t_x=0.000625$ sec. During the excitation interval, the flux density of core A is decreasing from positive saturation due to the voltage on its control winding which is the voltage from the control winding of core B minus the control voltage less the drop across the control resistance. This flux density change is defined as $\beta_{A,x,R}$ and is represented between points 2 and 3 of Fig. 2 and is determined by equation 5.

$$\beta_{A,x,R}=\frac{t_x e_{c,A}10^{-8}}{A_{Fe}N_c}=\frac{t_x 10^{-8}}{A_{Fe}N_c}\times$$
$$\left[E-E_c-\frac{H_{c,2}l_{Fe}R_c}{N_c}\right] \quad (5)$$

Applying the constants of the given amplifier, $\beta_{A,x,R}=96{,}000$ lines/in.^{2}. At the time core B is about to go into saturation, the load current is at point 2, Fig. 4(A); the control current is at point 2, Fig. 4(B); core B is at point 7, Fig. 2($\beta=+\beta_s, H=+H_{c,2}$); and core A is at point 3, Fig. 2 ($\beta=+4{,}000$, $H=-H_{c,2}$). Now core B saturates and its magnetic force is represented by point 8, Fig. 2. The load current ($i_{L,s}$) increases to a value given by equation 6 and is represented by point 3, Fig. 4(A). For the given amplifier $i_{L,s}=5.0$ amp.

$$i_{L,s}=\frac{E}{R_G+R_L} \quad (6)$$

Since there is no induction when core B is fully saturated, $e_{c,B}=0$. Now the only voltage in the control circuit is the relatively low control voltage, and the rate of change of flux density in core A decreases greatly. The assumption used in this analysis is that the coercive force of core A returns to the d-c hysteresis loop ($-H_{c,1}$) which is represented by point 4, Fig. 2. The control current at this time ($i_{c,s}$) is at point 3, Fig. 4(B), and is given by equation 7.

$$i_{c,s}=-\frac{l_{Fe}H_{c,1}}{N_c} \quad (7)$$

Using the given constants, $i_{c,s}=-0.02$ amp.

The load and control currents now remain at these values until the end of the saturation interval and are represented respectively at that time by point 4, Fig. 4(A) and point 4, Fig. 4(B). During the saturation interval ($T-t_x$), the control voltage causes a further decrease in flux density in core A. This is represented between points 4 and 5 of Fig. 2, which is defined as $\beta_{A,s,R}$ and is given by equation 8.

$$\beta_{A,s,R}=\frac{(T-t_x)e_{c,B}}{A_{Fe}N_c}10^8$$

$$=\frac{(T-t_x)10^8}{A_{Fe}N_c}\left[-E_c-\frac{H_{c,1}l_{Fe}R_c}{N_c}\right] \quad (8)$$

Again using the constants of the given amplifier in equation 8, $\beta_{A,s,R}=4{,}000$ lines/in.^{2}.

At the end of the half-cycle (point 4, Fig. 3), core B returns to point 1, Fig. 2 ($\beta=+\beta_s$, $H=-H_{c,1}$). Core A has returned from positive saturation flux density an amount equal to $\beta_{A,x,R}+\beta_{A,s,R}$ which in the given amplifier is 100,000 lines/in.^{2} and therefore core A is at point 5, Fig. 2 ($\beta=0$, $H=-H_{c,1}$). Thus, the conditions in the cores at the beginning of the positive half-cycle of e_A are exactly interchanged and thus fulfill the criterion for steady-state operation. The average control current I_c is given by equation 9.

$$I_c=\frac{1}{T}\int_0^{t_x} i_{c,x}dt+\frac{1}{T}\int_{t_x}^{T} i_{c,s}dt \quad (9)$$

If the excitation component of the load current is neglected, the average load current (I_L) is given by equation 10.

$$I_L=\frac{1}{T}\int_{t_x}^{T}\frac{E}{R_G+R_L}dt \quad (10)$$

Using the constants of the demonstration amplifier, the average control and load current for the mode investigated were respectively −0.04 and 2.5 amp.

The theoretical $\beta-H$ loop followed by a core in a full-cycle of operation, and for the given conditions, is outlined by the numbers 1 through 8 in Fig. 2.

General Analysis

The maximum voltage that can be applied to a magnetic amplifier and still maintain full control is given by equation 1. However, it may be desirable in some cases to design an underexcited magnetic amplifier in which the maximum flux density swing in a core is not $2\beta_s$ at the design voltage. Define a flux density β_s' such that $\beta_s'<\beta_s$. Also define a quantity as in equation 11.

$$f=\frac{1}{2T} \quad (11)$$

Now, eliminating T from equation 1 by means of equation 11 and replacing β by β_s', equation 12 is obtained which defines the supply voltage of the amplifier.

$$E=4fN_GA_{Fe}\beta_s'\times10^{-8} \quad (12)$$

Assume the amplifier is operating at a general steady-state operating point and time is such that a positive half cycle of voltage to the gate winding of core B is just beginning (point 1, Fig. 3). At this time, core A will be at point 1, Fig. 2 ($\beta=+\beta_s$, $H=-H_{c,1}$) and core B will be at some general level of flux density at a coercive force of $-H_{c,1}$. As the voltage e_B rises to E (point 2, Fig. 3) and neglecting the voltage drop over the gate and load resistance by the excitation current, the gate voltage of core B is $e_{G,B,x}=E$. The voltage across the control coil of core B at this time is given by equation 13.

$$e_{c,B,x}=\frac{N_c}{N_G}e_{G,B,x}=\frac{N_c}{N_G}E \quad (13)$$

The control voltage $e_{c,B,x}$, minus the control voltage E_c less the drop over the control resistance drives core A to a coercive force of $-H_{c,2}$ and a control current flows equal to:

$$i_{c,x}=-\frac{H_{c,2}l_{Fe}}{N_c} \quad (14)$$

The current in the gate winding of core B and load at this time is that required through the gate turns to overcome the coercive force of core B plus the reflected control current. This is given in equation 15.

$$i_{G,B,x}=i_{L,x}=\frac{H_{c,2}l_{Fe}}{N_G}-\frac{N_c}{N_G}i_{c,x}=\frac{2H_{c,2}l_{Fe}}{N_G} \quad (15)$$

Now as time progresses, the flux density of core B will increase toward positive saturation. If the flux density change required to produce saturation is defined as $\beta_{B,x,F}$, the time for saturation to occur or the excitation interval (t_x) is given by equation 16.

$$t_x=\frac{N_GA_{Fe}\beta_{B,x,F}}{E}10^{-8} \quad (16)$$

The flux density change of core A from positive saturation during the excitation interval is defined as $\beta_{A,x,R}$. This flux density change is due to the voltage impressed upon the control coil of core A and is given by equation 17.

$$\beta_{A,x,R}=\frac{10^8}{N_cA_{Fe}}\left[\frac{N_c}{N_G}E-E_c-\frac{H_{c,2}l_{Fe}R_c}{N_c}\right]t_x \quad (17)$$

After core B saturates, the load current increases to a value limited by the load and gate resistance. The induced voltage from the control coil of core B becomes zero and the voltage in the control circuit reduces to the control voltage (E_c). It is assumed that the control voltage is small enough that core A returns to the coercive force of the d-c hysteresis loop ($-H_{c,1}$). During the saturation interval the control voltage causes an additional change of flux density in core A from positive saturation defined as $\beta_{A,s,R}$ and given by equation 18.

$$\beta_{A,s,R}=\frac{10^8}{N_cA_{Fe}}\left[-E_c-\frac{H_{c,1}l_{Fe}R_c}{N_c}\right](T-t_x) \quad (18)$$

The criterion for steady-state operation as stated previously is given by

$$\beta_{B,x,F}=\beta_{A,x,F}+\beta_{A,s,R} \quad (19)$$

Substituting from equations 16, 17, and 18 into 19, the following relation is obtained:

$$E_c=-\frac{l_{Fe}R_c(H_{c,2}-H_{c,1})}{N_c}\frac{t_x}{T}-\frac{l_{Fe}R_cH_{c,1}}{N_c} \quad (20)$$

where

$$-\frac{l_{Fe}R_cH_{c,2}}{N_c}\lesseqgtr E_c\lesseqgtr\frac{l_{Fe}R_cH_{c,1}}{N_c}$$

The average load current as a function of the duration of the excitation interval is given by equation 21.

$$I_L=\frac{1}{T}\int_0^{t_x} i_{L,x}dt+\frac{1}{T}\int_{t_x}^{T}\frac{E}{R_G+R_L}dt \quad (21)$$

Substituting for $i_{L,x}$ from equation 15 into 21 and integrating:

$$I_L=\frac{2H_{c,2}l_{Fe}}{N_G}\frac{t_x}{T}+\frac{E}{R_L+R_G}\left(\frac{T-t_x}{T}\right) \quad (22)$$

where

$$0\lesseqgtr t_x\lesseqgtr T$$

By combining equations 20 and 22, the steady-state control characteristic of average load current versus control voltage can be obtained as in equation 23.

$$I_L=\frac{2H_{c,2}}{N_G(H_{c,2}-H_{c,1})}\left[-l_{Fe}H_{c,1}-\frac{E_cN_c}{R_c}\right]+\frac{E}{(R_L+R_G)l_{Fe}(H_{c,2}-H_{c,1})}\left[l_{Fe}H_{c,2}+\frac{E_cN_c}{R_c}\right] \quad (23)$$

where

$$-\frac{H_{c,2}l_{Fe}R_c}{N_c}\lesseqgtr E_c\lesseqgtr-\frac{H_{c,1}l_{Fe}R_c}{N_c}$$

Since the characteristics of magnetic amplifiers are in general plotted in terms of load and control current, the control voltage in equation 23 may be replaced by control current as follows:

$$I_L=\frac{2H_{c,2}}{N_G(H_{c,2}-H_{c,1})}[-l_{Fe}H_{c,1}-N_cI_c]+\frac{E}{(R_L+R_G)l_{Fe}(H_{c,2}-H_{c,1})}\times[l_{Fe}H_{c,2}+N_cI_c] \quad (24)$$

where

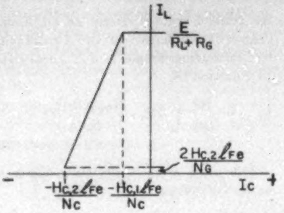


Fig. 5. Steady-state transfer function of amplifier

$$-\frac{H_{c,2}l_{Fe}}{N_c} \gtreqless I_c \gtreqless -\frac{H_{c,1}l_{Fe}}{N_c}$$

From equation 24 the steady-state amplifier characteristic may be plotted as in Fig. 5.

Equation 24 may be greatly simplified if the excitation component of load current is neglected. The effect of this is to drop the lower end of the characteristic, shown in Fig. 5, to zero. Using this simplification, equation 24 reduces as in 25.

$$I_L = \frac{E}{R_L+R_G}\left[\frac{H_{c,2}}{H_{c,2}-H_{c,1}} + \frac{N_c I_c}{l_{Fe}(H_{c,2}-H_{c,1})}\right] \tag{25}$$

where

$$-\frac{H_{c,2}l_{Fe}}{N_c} \gtreqless I_c \gtreqless \frac{H_{c,1}l_{Fe}}{N_c}$$

TRANSIENT RESPONSE ASSUMING NO RECTIFIER LEAKAGE

When the magnetic amplifier is operating steady-state, the control voltage must be such that the flux density change of the core going into positive saturation is equal to the total reset flux density change of the other core (equation 19). When this condition is transiently upset by a change in control voltage, the amplifier will seek a new period of excitation interval such that this condition will again prevail. The time for this to occur is of vital interest to the designers of magnetic amplifiers.

In studying the response of the amplifier, the cores need no longer be referred to as *A* and *B*. It is evident at this point that in any half-cycle, one core is going into positive saturation while the other is being reset from positive saturation. The core going into positive saturation determines the excitation interval of that half-cycle according to equation 16. If the flux density change required before saturation occurs is defined as β_F, then equation 16 can be expressed as in 26.

$$t_x = \frac{N_G A_{Fe} 10^{-8}}{E}\beta_F \tag{26}$$

The reset flux density change from positive saturation is defined as β_R and is obtained by adding equations 17 and 18 as follows:

$$\beta_R = \frac{10^8}{N_c A_{Fe}}\left[\frac{N_c}{N_G}E + \frac{l_{Fe}R_c}{N_c}(H_{c,1}-H_{c,2})\right]t_x + \frac{10^8}{N_c A_{Fe}}\left[-E_c - \frac{H_{c,1}l_{Fe}R_c}{N_c}\right]T \tag{27}$$

Now define

$$K_1 = \frac{N_G A_{Fe} 10^{-8}}{E} \tag{28}$$

$$K_2 = \frac{10^8}{N_c A_{Fe}}\left[\frac{N_c}{N_G}E + \frac{l_{Fe}}{N_c}(H_{c,1}-H_{c,2})\right] \tag{29}$$

$$K_c = \frac{10^8}{N_c A_{Fe}}\left[-E_c - \frac{H_{c,1}l_{Fe}R_c}{N_c}\right]T \tag{30}$$

K_c will be referred to as the control constant which may take on different values as a function of the control voltage E_c. Substituting from equations 28, 29, and 30 into 26 and 27 gives the following

$$t_x = K_1\beta_F \tag{31}$$

$$\beta_R = K_2 t_x + K_c \tag{32}$$

The criterion for steady-state operation is that $\beta_F = \beta_R$. If the steady-state excitation interval is defined as $t_{x,ss}$, equation 33 can be obtained from equations 31 and 32.

$$t_{x,ss} = \frac{K_1 K_c}{1-K_1K_2} \tag{33}$$

It is clear that the reset flux density of a core in one half-cycle becomes the forward flux density and determines the excitation interval in the following half-cycle. Keeping this in mind and by the use of equations 31 and 32, the transient response of the magnetic amplifier can be obtained.

Refer to Fig. 6 and assume that prior to the half-cycle represented by the shaded half-cycle the amplifier was at a steady-state condition determined by its control constant being at some specific value $K_c = K_{c1}$. The corresponding steady-state excitation is $t_{x,ss} = t_{x,ss1}$. From equation 33:

$$t_{x,ss1} = \frac{K_1 K_{c1}}{1-K_1K_2} \tag{34}$$

At the beginning of the shaded half-cycle, assume the control constant is changed by means of the control voltage to a new value $K_c = K_{c,2}$ and the response of the amplifier is desired in the following half-cycles: 1, 2, 3, ...*n*.

In the first half-cycle after the shaded one, the excitation time will be determined by the core going into positive saturation according to equation 31. Thus:

$$t_{x,1} = K_1\beta_{F1} \tag{35}$$

However, the forward flux density in the first half-cycle was the reset flux density of the shaded half-cycle which was from equation 32:

$$\beta_{R0} = K_2 t_{x,ss1} + K_{c,2} \tag{36}$$

Since $\beta_{F1} = \beta_{F0}$, using 35 and 36 gives:

$$t_{x,1} = K_1[K_2 t_{x,ss1} + K_{c,2}] \tag{37}$$

The reset flux density in the first half-cycle according to equation 32 is given by equation 38.

$$\beta_{R1} = K_2 t_{x,1} + K_{c,2} \tag{38}$$

Combining equations 37 and 38 gives:

$$\beta_{R1} = (K_1K_2^2)t_{x,ss1} + K_1K_2K_{c,2} + K_{c,2} \tag{39}$$

In the second half-cycle, the excitation interval $t_{x,2}$ is:

$$t_{x,2} = K_1\beta_{F2} = K_1\beta_{R1} \tag{40}$$

Substituting for β_{R1} from equation 39 gives:

$$t_{x,2} = (K_1K_2)^2 t_{x,ss1} + K_1^2K_2K_{c,2} + K_1K_{c,2} \tag{41}$$

This process may be continued to the *n*th half-cycle and the excitation interval for this half-cycle may be expressed as:

$$t_{x,n} = (K_1K_2)^n t_{x,ss1} + \frac{K_{c,2}}{K_2}[K_1K_2 + (K_1K_2)^2 + \ldots + (K_1K_2)^n] \tag{42}$$

Equation 42 may be rewritten as follows:

$$t_{x,n} = (K_1K_2)^n t_{x,ss1} + K_1K_{c,2}[1 + K_1K_2 + (K_1K_2)^2 + \ldots + (K_1K_2)^{n-1}] \tag{43}$$

Summing the geometric series of equation 43, it may be expressed as follows:

$$t_{x,n} = (K_1K_2)^n t_{x,ss1} + \frac{K_1K_{c,2}}{1-K_1K_2}[1-(K_1K_2)^n] \tag{44}$$

Now by the use of equation 33, the ultimate steady-state excitation interval $t_{x,ss2}$ may be expressed as in equation 45.

$$t_{x,ss2} = \frac{K_1K_{c,2}}{1-K_1K_2} \tag{45}$$

Substituting from equation 45 into 44 gives:

$$t_{x,n} = (t_{x,ss1} - t_{x,ss2})(K_1K_2)^n + t_{x,ss2} \tag{46}$$

By use of equation 46, the response of the magnetic amplifier can be determined

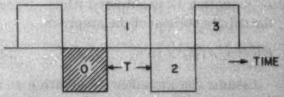


Fig. 6. Square-wave supply voltage

regardless of its speed of response. Such an expression is useful where response times approach the half-cycle time.

If the response time of the amplifier is relatively long compared to a half-cycle a time constant may be derived. The following equivalency may be written as follows:

$$(K_1K_2)^n = \epsilon^{n \log_e K_1K_2} \tag{47}$$

Equations 46 and 47 can be combined as in 48:

$$t_{x,n} = (t_{x,s.1} - t_{x,ss2})\epsilon^{n \log_e K_1K_2} + t_{x,ss2} \tag{48}$$

Since the response time is assumed to be long compared to half-cycle time, time as a function of the nth half-cycle and half-cycle time is approximately:

$$t \cong nT \tag{49}$$

Combining equations 48 and 49:

$$t_x = (t_{x,ss1} - t_{x,ss2})\epsilon^{-\left(\frac{\log_e K_1K_2}{T}\right)t} + t_{x,ss2} \tag{50}$$

It is evident from equation 50 that the response of the amplifier is exponential. The fact that $K_1K_2<1$ will be apparent later. Thus, the term $\log_e K_1K_2<0$ and since time constants are normally expressed as positive numbers, the negative exponent was introduced in equation 50. Define the term "time constant" (T_c) as follows:

$$T_c = -\frac{T}{\log_e K_1K_2} \tag{51}$$

Using equations 28 and 29, K_1K_2 can be expressed as follows:

$$K_1K_2 = 1 - \frac{N_G l_{Fe} R_c}{N_c^2 E}(H_{c,2} - H_{c,1}) \tag{52}$$

Eliminate E from equation 52 by means of equation 12 and define a quantity x as in equation 53.

$$K_1K_2 = 1 - \frac{l_{Fe} R_c 10^8}{4fN_c^2 A_{Fe} \beta_s'}(H_{c,2} - H_{c,1}) = 1 - x \tag{53}$$

Now eliminate K_1K_2 from equation 51 by means of the quantity defined in equation 53.

$$T_c = -\frac{T}{\log_e (1-x)} \tag{54}$$

Restricted to $0 \gtreqless x << 1$

It is obvious from equation 54 that $0 \gtreqless x \gtreqless 1$ covers a range of time constants from half-cycle time T to infinity. A previous assumption presumed investigating only response times that were relatively long compared to half-cycle time, therefore the range of x is restricted to $0 \gtreqless x << 1$.

Following is the infinite series expression for $\log_e(1-x)$.

$$\log_e(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots\ldots \tag{55}$$

An approximation for equation 55 when $x<<1$ is given in equation 56.

$$\log_e(1-x) \cong -x \tag{56}$$

when $x<<1$

Now substitute for $\log_e(1-x)$ from equation 56 into 54 and replace x by the value defined in equation 53.

$$T_c = \frac{2N_c^2 A_{Fe} \beta_s' 10^{-8}}{l_{Fe} R_c (H_{c,2} - H_{c,1})} \tag{57}$$

POWER GAIN (INFINITE RECTIFIER LEAKAGE RESISTANCE)

Power gain P_G is a useful measure of amplifier performance and is defined as follows:

$$P_G = \left(\frac{dI_L}{dI_c}\right)^2 \frac{R_L}{R_c} \tag{58}$$

Using equations 12, 25, and 58, the following expression for power gain can be obtained:

$$P_G = \frac{16f^2 N_G^2 N_c^2 A_{Fe}^2 \beta_s'^2 10^{-16}}{\frac{R_c}{R_L}(R_L + R_G)^2 l_{Fe}^2 (H_{c,2} - H_{c,1})^2} \tag{59}$$

DYNAMIC POWER GAIN (INFINITE RECTIFIER LEAKAGE RESISTANCE)

Dynamic power gain is defined as:

$$P_D = \frac{\left(\frac{dI_L}{dI_c}\right)^2 \frac{R_L}{R_c}}{T_c} = \frac{P_G}{T_c} \tag{60}$$

Substituting from equation 57 and 59 into 60:

$$P_D = \frac{8f^2 N_G^2 R_L A_{Fe} \beta_s' 10^{-8}}{(R_L + R_G)^2 l_{Fe} (H_{c,2} - H_{c,1})} \tag{61}$$

USE OF EQUATIONS

Previously, constants were assigned to a demonstration amplifier. If these constants are applied to equations 25, 57, 59 and 61, the results are as follows:

$I_L = 5[1.5 + 25I_c]$

where $-0.06 \gtreqless I_c \gtreqless -0.02$

$T_c = 0.0156$ sec.
$P_G = 1170$
$P_D = 75{,}000$

STEADY-STATE CHARACTERISTICS ASSUMING RECTIFIER LEAKAGE

The switches of Fig. 1 are now closed thus simulating rectifier leakage R_l. The excitation interval is still described by equation 16. If it is assumed that the voltage drop across the control resistance is small compared to induced voltages in the control winding, then $e_{c,B} \cong e_{c,A}$ and $e_{G,B} \cong e_{G,A}$ If this is true, the voltages $e_{G,A} \cong e_A$ and $e_{G,B} \cong e_B$ and the leakage resistance will have little influence on the reset flux density during the excitation interval and equation 17 still describes the reset flux density change during this interval. During the saturation interval, the leakage resistance will have a pronounced influence on the reset flux density. If gate and forward rectifier resistance are assumed negligible compared to load resistance, it can be shown that the reset flux density change in this interval is described by equation 62.

$$\beta_{A,S,R} = \frac{10^8}{N_c A_{Fe}\left[1 + \left(\frac{N_G}{N_c}\right)^2 \frac{R_c}{R_l}\right]} \times \left[-E_c - \frac{H_{c,1} l_{Fe} R_c}{N_c} + \frac{2EN_G R_c}{N_c R_l}\right](T - t_x) \tag{62}$$

To simplify the analysis, it will be assumed that:

$$1 >> \left(\frac{N_G}{N_c}\right)^2 \frac{R_c}{R_l}$$

This assumption implies that the leakage current is derived from a constant current source or that the leakage resistance is very large compared to the reflected control resistance. This condition normally exists except where deliberate leakage is introduced as a negative feedback. With this assumption, equation 62 reduces to:

$$\beta_{A,S,R} = \frac{10^8}{N_c A_{Fe}}\left[-E_c - \frac{H_{c,1} l_{Fe} R_c}{N_c} + \frac{2EN_G R_c}{N_c R_l}\right](T - t_x) \tag{63}$$

The conditions that must exist for steady-state operation are still described by equation 19. Substituting into equation 19 from equations 16, 17, and 63 results in equation 64.

$$I_c = -\frac{t_x}{T}\frac{l_{Fe}}{N_c}\left[H_{c,2} - H_{c,1} + \frac{2EN_G}{R_l l_{Fe}}\right] - \frac{l_{Fe}}{N_c}\left[H_{c,1} - \frac{2EN_G}{l_{Fe} R_l}\right] \tag{64}$$

Using equation 22 and 64 and neglecting the excitation component of load current, the following expression for load current can be obtained as in equation 65

$$I_L = \frac{E}{R_L + R_G}\left[\frac{H_{c,2}}{H_{c,2} - H_{c,1} + \frac{2EN_G}{R_l l_{Fe}}} + \frac{N_c I_c}{l_{Fe}\left[H_{c,2} - H_{c,1} + \frac{2EN_G}{R_l l_{Fe}}\right]}\right] \tag{65}$$

where

$$-\frac{l_{Fe} H_{c,2}}{N_c} \gtreqless I_c \gtreqless -\frac{l_{Fe}}{N_c}\left(H_{c,1} - \frac{2EN_G}{l_{Fe} R_l}\right)$$

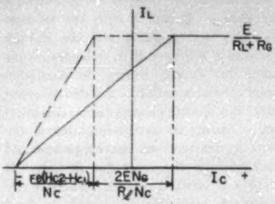


Fig. 7. Steady-state transfer function with rectifier leakage

Load current as a function of control current is plotted in Fig. 7. The dashed curve indicates the characteristic with infinite rectifier leakage resistance.

Transient Response with Rectifier Leakage

The transient response may be analyzed in exactly the same manner as outlined previously. Equation 28 is unchanged and 29 and 30 becomes respectively:

$$K_2=\frac{10^8}{N_cA_{Fe}}\left[\frac{N_c}{N_G}E+\frac{l_{Fe}R_c}{N_c}(H_{c,1}-H_{c,2})-\frac{2EN_GR_c}{N_cR_l}\right] \quad (66)$$

$$K_c=\frac{10^8}{N_cA_{Fe}}\left[-E_c-\frac{H_{c,1}l_{Fe}R_c}{N_c}+\frac{2EN_GR_c}{N_cR_l}\right] \quad (67)$$

Following the procedure established previously, it can be ascertained that the time constant including rectifier leakage ($T_{c,l}$) is expressed by equation 68.

$$T_{c,l}=\frac{T}{\frac{l_{Fe}R_c10^8}{4fN_c^2A_{Fe}\beta_s'}(H_{c,2}-H_{c,1})+\frac{2N_G^2R_c}{N_c^2R_l}} \quad (68)$$

The expression for time constant without rectifier leakage (T_c), given by equation 57, can be substituted in equation 68 which can then be rewritten as follows:

$$T_{c,l}=\frac{T_c}{1+\frac{4fN_G^2R_c}{N_c^2R_l}T_c'} \quad (69)$$

Investigation discloses that rectifier leakage is a type of negative feedback. Further analysis will show that decreasing rectifier leakage resistance will reduce the dynamic power gain, thus resulting in a longer time constant for a given power gain in a specific amplifier. In general, speed of response may be traded for stability of the transfer characteristic by means of negative feedback. However, when the negative feedback is introduced by means of rectifier leakage, the feedback function will be variable if leakage is a function of temperature. Thus, the over-all effect of rectifier leakage is detrimental.

Qualitative Discussion of Positive and Negative Feedback

Effects of positive and negative feedback may be analyzed by the method outlined previously. If a negative feedback is derived by passing the load current of the amplifier through a small control winding, it can generally be assumed that the feedback is from a constant current source to the extent that transient voltages induced into the load circuit through the feedback winding may be neglected. Since the feedback is only active during the saturation interval, only the reset flux density equation during the saturation interval is altered. Equation 62 can be rewritten to include the effect of positive and negative current feedback as follows:

$$\beta_{A,S,R}=\frac{10^8}{N_cA_{Fe}}\left[-E_c-\frac{H_{c,1}l_{Fe}R_c}{N_c}+\frac{2EN_GR_c}{N_cR_l}\pm\frac{EN_fR_c}{N_c(R_L+R_G)}\right](T-t_x) \quad (70)$$

Here the positive and negative signs are associated with negative and positive feedback respectively.

Again the procedures for obtaining response, power gain, etc. are as outlined previously. With negative feedback, it will be found that the dynamic power gain is lowered. The loss in speed of response, however, has been traded for stability of the transfer characteristic. With positive feedback, the converse is true.

Steady-State Transfer Characteristic with Sinusoid Supply

If the assumption is made that during the excitation interval a core is being reset along the dynamic hysteresis loop and during the saturation interval it is being reset along the d-c loop, equation 20 is valid for a sine wave of supply voltage to the gate windings. Equation 20 may be rewritten in terms of control current and a "firing" angle α_x as follows:

$$I_c=-\frac{l_{Fe}(H_{c,2}-H_{c,1})}{N_c}\frac{\alpha_x}{\pi}-\frac{H_{c,1}l_{Fe}}{N_c} \quad (71)$$

where $0\lesseqgtr\alpha_x\lesseqgtr\pi$. The average load current as a function of firing angle is given by equation 72:

$$I_L=\frac{1}{T}\left(\frac{E_m}{R_L+R_G}\right)\int_{\alpha x/\omega}^{\pi/\omega}\sin\omega t\,dt \quad (72)$$

Performing the indicated integration:

$$I_L=\frac{E_m}{\pi(R_L+R_G)}[1+\cos\alpha_x] \quad (73)$$

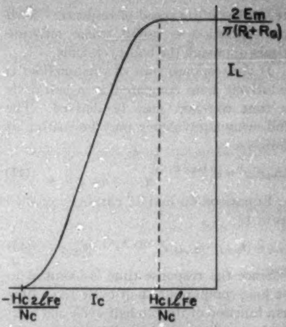


Fig. 8. Theoretical transfer characteristic with sinusoidal supply voltage

Average load current as a function of average control current may be obtained by eliminating α_x from equations 71 and 73 as follows:

$$I_L=-\frac{E_m}{\pi(R_L+R_G)}\left[1+\cos\frac{-N_c\pi}{l_{Fe}(H_{c,2}-H_{c,1})}\times\left(I_c+\frac{H_{c,1}l_{Fe}}{N_c}\right)\right] \quad (74)$$

The theoretical steady-state transfer function of a magnetic amplifier with sinusoidal supply voltage is shown in Fig. 8.

Test Results

A toroidal amplifier using Orthonol cores was constructed and tested in the circuit of Fig. 1 and with a square-wave supply voltage. The constants of the amplifier were as follows:

$N_G=535$ turns
$N_c=600$ turns

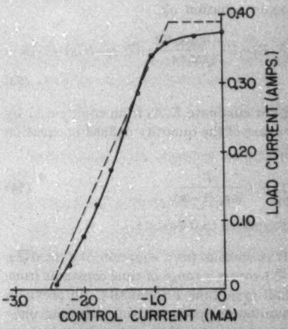


Fig. 9. Test amplifier transfer characteristic with square wave supply voltage

——Actual
– – –Theoretical

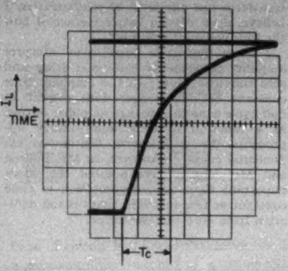


Fig. 10 (left). Response time of test amplifier (sweep time 20 milliseconds/centimeter)

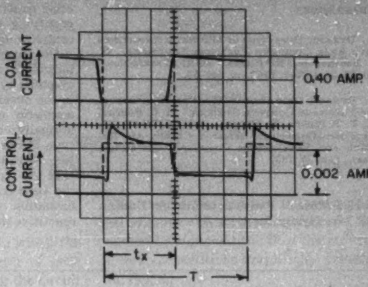


Fig. 11 (right). Load and control current of test amplifier

$E=40$ volts
$T=1/800$ sec.
$H_{c,1}=0.114$ AT/in.
$H_{c,2}=0.342$ AT/in.
$l_{Fe}=3.94$ in.
$A_{Fe}=0.084$ in.^{2}
$R_G=3$ ohms
$R_L=100$ ohms
$R_c=1000$ ohms

Using the foregoing constants, the transfer characteristic of the amplifier according to equation 25 is:

$$I_L=0.388[1.5+666I_c]$$

where $-0.0025 \gtreqless I_c \gtreqless -0.00075$

The theoretical amplifier characteristic is plotted as a dashed line on Fig. 9 and the actual test data are the solid line curve.

The amplifier time constant calculated by means of equation 57 is as follows:

$$T_c=\frac{2(600)^2(0.084)(55{,}500)10^{-8}}{3.94(1{,}000)(0.342-0.114)}=0.037 \text{ sec}$$

Fig. 10 shows the actual response of the amplifier as a function of time on a cathode-ray oscillograph and recorded by a Land camera. The sweep calibration on the cathode-ray oscillograph was 20 milliseconds per centimeter. The measured time constant was 0.039 sec which is in close agreement with the calculated value.

Using the constants of the test amplifier, the instantaneous load and control currents during the excitation and saturation interval may be calculated as follows:

$$i_{c,s}=-\frac{l_{Fe}H_{c,1}}{N_c}=-\frac{3.94(0.114)}{600}$$
$$=-0.00075 \text{ amp}$$

$$i_{c,x}=-\frac{l_{Fe}H_{c,2}}{N_c}=-\frac{3.94(0.342)}{600}$$
$$=-0.00225 \text{ amp}$$

$$i_{L,s}=\frac{E}{R_L+R_G}=\frac{40}{103}=0.388 \text{ amp}$$

Fig. 11 shows the actual instantaneous load and control currents during the excitation and saturation interval. The dashed lines in Fig. 11 show the theoretical values of these currents to be in close agreement with actual values.

Conclusion

A method of calculating the performance of an amplifier utilizing square-loop core material an alternating square-wave voltage supply has been shown to be in close agreement with observed data. It is felt that the elimination of trigonometric functions allows a clearer physical picture of how the amplifier operates. The theoretical steady-state transfer function was extended to the case of a sinusoidal supply voltage. The time constant derived for the square wave of supply voltage indicates that the time constant is proportional to the supply voltage. It is felt that if this analysis were extended to the general supply voltage, the time constant at any point on the transfer function might be found to be proportional to the slope of the transfer function.

References

1. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE-LOOP CORE MATERIALS, H. F. Storm. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 629–40.

2. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

3. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

◆

Discussion

H. C. Bourne, Jr., and D. Nitzan (University of California, Berkeley, Calif.): The discussers note that the author has used the same model for the rectangular core material and consequently obtains the identical results for the center-tap circuit with no rectifier leakage as obtained in a previous paper[1] for the alternating-output circuit. Rectifier leakage, when described in terms of a reverse rectifier resistance, produces slightly different results for the three basic self-saturating circuits because of the different reverse voltages involved. In a less available publication[2] a general analysis of all three circuits with square-wave supply voltage and rectifier leakage is in complete agreement with the author's results.

The author uses a repetitive technique to obtain an expression for the excitation time in the *n*th half-cycle, reduces the series form to a closed form, and obtains an expression for the time constant. It might be pointed out that the implied difference equation

$$t_{x,n+1}=K_1B_{F,n+1} \tag{31A}$$

$$B_{R,n}=K_2t_{x,n}+K_{c,n} \tag{32A}$$

$B_{R,n}=B_{F,n+1}$ (from statement following equation 33)

$$t_{x,n+1}=K_1K_2t_{x,n}+K_1K_{c,n}$$

may be transformed directly using Laplace or *z*-transform techniques to obtain the same result in the form of a dynamic transfer function. Possibly the advantage of the series form lies in the clear physical picture which it gives of the discrete nature of the magnetic amplifier operation in the real time domain. However, from a systems viewpoint, the ultimate goal is a dynamic transfer function, which can be constructed from the author's results, of course.

The discussers believe that the author has been too severe, at least from a mathematical viewpoint, in the restrictions placed on the expressions for time constant in equations 50 and 54. Certainly, when using the term "time constant" in dealing with discrete systems some care in interpretation is necessary, but it seems perfectly logical to interpret $K_1K_2=0$, $x=1$, $T_c=0$ as an exponential response with zero time constant. Since in the author's definition, time did not start until a half-cycle after the application of the change in control voltage, the familiar half-cycle delay has been correctly implied. The actual restrictions on the validity of these expressions for fast response times concern the validity of the core representation and the possibility of gate rectifier unblocking under these conditions.

REFERENCES

1. DYNAMIC OPERATION OF MAGNETIC AMPLIFIERS WITH REAL CORE AND RECTIFIER FUNCTIONS, H. C. Bourne, Jr., D. Nitzan. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 646–55.

2. DYNAMIC OPERATION OF MAGNETIC AMPLIFIERS FOR FEEDBACK CONTROL SYSTEMS, H. C. Bourne, Jr., T. T. Kadota, D. Nitzan. *AFCRC-TR-57-167*, Air Force Cambridge Research Center, Air Research and Development Command, Cambridge, Mass., Feb. 28, 1957.

H. H. Britten: I wish to thank Mr. Bourne and Mr. Nitzan for their references to two other papers of similar content. It is indeed gratifying that independent works are in complete agreement.

One purpose of this paper was to help formulate a physical picture of magnetic amplifier type of operation. This is needed in the development of new circuits that use magnetic amplifiers outside of what is normally considered magnetic amplifier type of operation. Thus, Mr. Bourne and Mr. Nitzan are correct in their assumption that the repetitive technique, for obtaining the time constant, was used to aid in the formation of the physical picture. The repetitive method is not as concise as the *z*-transform but the level of mathematics, I believe, does allow a better picture of how the time constant is formed.

I agree that it is possible to interpret amplifier response as half-cycle delay and zero time constant. As the amplifier time constant becomes longer with respect to half-cycle time the average amplifier output as a function of time more nearly approaches the well known continuous exponential curve. However, as Mr. Bourne and Mr. Nitzan have pointed out, more important than the interpretation of time constant at fast response times is the deviation from core representation.

Triggering in the Reversible Half-Wave Magnetic Amplifier

D. L. MCMURTRIE
MEMBER AIEE

J. P. WARD
ASSOCIATE MEMBER AIEE

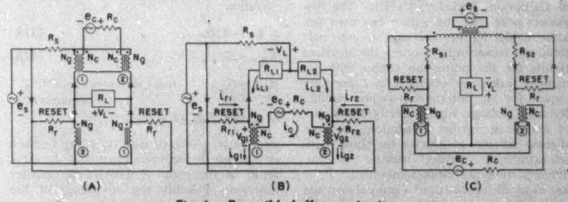
THIS PAPER describes a type of instability or triggering observed in the half-wave bridge circuit. The circuit, described in the literature by Lufcy and others, is shown schematically in Fig. 1(A).[2–4] An explanation of this triggering is presented based upon a type of reverse-slope magnetic-loop operation not previously described in the magnetic-amplifier literature (see acknowledgment). Figs. 2(A) and (B) are two commonly used models of the dynamic loops of "square-loop" magnetic-core materials. The paper shows that neither model predicts the observed triggering but that a loop having the "reverse slope" of Fig. 2(C) readily predicts the instability. In Fig. 2(C) the lower minor loop lies to the right of upper minor loop, and lines connecting their centers or flanks would have a negative or reverse slope. The reality of this reverse-slope phenomenon has been determined experimentally, and verifying data are presented. The reader is cautioned that the reverse-slope phenomenon applies only to minor hysteresis loops and not to major loops operating over the entire range from positive to negative saturation. The reader should also be aware that no model of minor- or major-loop operation can be expected to apply to other than a limited range of exciting conditions.

Nomenclature

e_c, E_{cm}, E_c = instantaneous, peak, and rms control voltage
e_s, E_{sm}, E_s = instantaneous, peak, and rms supply voltage
i_c = control current
i_{L1}, i_{L2} = load currents in R_{L1} and R_{L2}
i_{G1}, i_{G2} = currents in gate windings of either core
i_{r1}, i_{r2} = currents in the reset circuit
N_c = control turns
N_g = gate or load turns
N_m = pickoff turns for induced voltage measurement
$(Ni)_1$, $(Ni)_2$ = total magnetic intensity applied to either core, ampere-turns: $(Ni)_1 = -N_c i_c + N_g i_{g1}$, $(Ni)_2 = N_c i_c + N_g i_{g2}$
R_c = control circuit resistance
R_{L1}, R_{L2} = load resistors for case of split load
R_L = load resistor $= R_{L1} + R_{L2}$ in case of split load
R_{mc}, R_{mr}, R_{ml} = measuring resistors in control, reset, and load circuits respectively (Fig. 10)
R_{r1}, R_{r2} = reset resistors
R_S = resistance in series with bridge or output circuit
R_{wc} = winding resistance of N_c
R_{wg} = winding resistance of N_g
V_{g1}, V_{g2} = voltage induced in gate windings N_g of either core
V_L, $\bar{V}_L$ = instantaneous and average load voltage across R_L or R_{L1} and R_{L2} combined
θ_1, θ_2 = firing or saturation angles of cores *1* and *2*
ψ = core flux
RHWACDC = reversible half-wave a-c, d-c
RHWAC = reversible half-wave a-c
RHWDC = reversible half-wave d-c

Reversible Half-Wave Magnetic Amplifier, Bridge and Center-Tap Types

The circuit being dealt with is the half-wave bridge circuit and its equivalent center-tap circuits shown in Fig. 1. This



(A) (B) (C)

Fig. 1. Reversible half-wave circuits

A—Bridge circuit
B—Equivalent center-tap-load circuit
C—Equivalent center-tap-supply circuit

Paper 58-776, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 11, 1958; made available for printing April 1, 1958.

D. L. MCMURTRIE and J. P. WARD are with the Raytheon Manufacturing Company, Wayland, Mass.

As the manuscript of this paper was in final preparation, the authors encountered a paper by L. A. Finzi and D. L. Critchlow.[1] As part of a general treatment of the reset process under various conditions, their paper describes magnetic-minor-loop operation similar to that reported in this paper and connects this loop operation with certain types of amplifier triggering. Happily, the amplifiers dealt with and the details of the triggering are not the same as those of this paper. The authors are pleased that this paper has come to their attention and wish to give due credit to the fine work of F. A. Finzi and D. L. Critchlow.

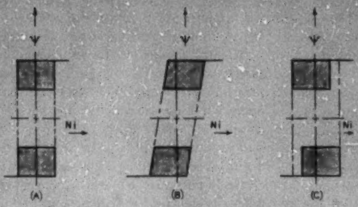


Fig. 2 (left). Magnetic - core models

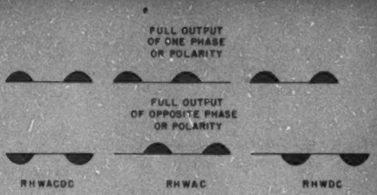

FULL OUTPUT OF ONE PHASE OR POLARITY
FULL OUTPUT OF OPPOSITE PHASE OR POLARITY
RHWACDC
RHWAC
RHWDC

Fig. 3 (right). Different types of reversible half-wave output

type of circuit has found extensive use as a servoamplifier because of its reversible output and fast response.* Operation of the circuit has been adequately described in the references cited previously, and the description given here is very brief.

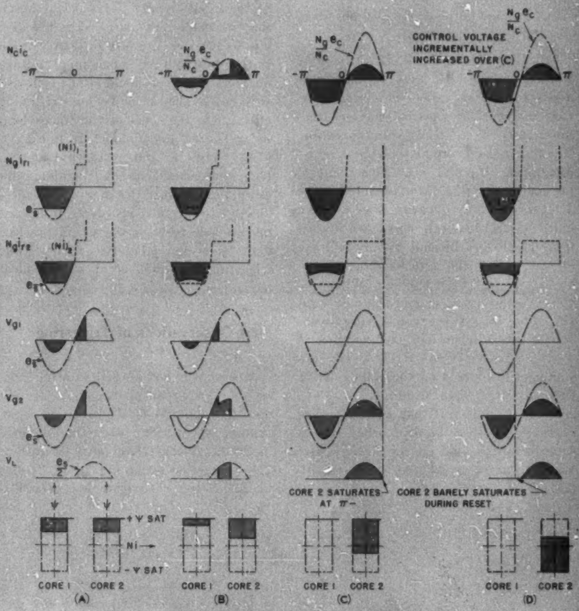
The waveforms shown in Fig. 4 indicate the operation of the circuit of Fig. 1(B) at various output levels assuming ideal rectangular-loop operation, ideal rectifiers, and zero winding resistance. It is also assumed that the control circuit resistance is matched to the reset resistance such that $R_c(N_g/N_c)^2=2R_r$. The waveforms of V_L show the nature of the output pulses which have d-c and fundamental frequency components (and higher harmonics).* Pulses of the opposite polarity would be produced if the phase of the control signal were reversed causing cores *1* and *2* to change places. In all cases, the negative or reset portions of coil voltages V_{g1} and V_{g2} are dictated by combined reset and control circuit action, the load circuit being kept open-circuited by its rectifiers. Positive or prefire voltage is dictated by the load circuit, the reset circuit being open-circuited by its rectifiers, and the control circuit having little effect because of its high impedance relative to the load circuit: i.e., $R_s,\ R_{L1},\ R_{L2} << R_r,\ R_c(N_g/N_c)^2$. For zero input and output the reset resistors are adjusted to cause approximately two thirds of the supply voltage area to appear across each resistor and one third across each gate winding during reset. The cores then saturate at approximately 50 degrees on the following positive half-cycle.

Determination of the instantaneous values of the control and reset current and voltage at finite levels of input and output becomes somewhat involved algebraically although the circuit is piecewise linear and the circuit equations not difficult to set up. The way in which the control voltage affects the reset of both cores is best seen by redrawing the circuit of Fig. 1(B) to show only those portions normally active during the reset half-cycle. This is done in Fig. 5. The three loop equations can readily be written remembering that in addition to the Kirchhoff equations there are also the $\psi_{gs}Ni$ or $V_{gs}i$ equations of the reactors (graphically presented by the loop paths).

The reset equations are not set up and solved, but the following observations should be sufficient to allow the reader to check the solutions graphically presented in Fig. 4:

1. The voltage and current equations are

$$V_{g1}=e_s-i_{r1}R_{r1} \qquad R_{r1}=R_{r2}=R_r$$
$$V_{g2}=e_s-i_{r2}R_{r2}$$
$$V_{g2}-V_{g1}=(e_c-i_cR_c)(N_g/N_c)$$
$$\left.\begin{array}{l} N_gi_{r1}-N_ci_c=(Ni)_1 \\ N_gi_{r2}+N_ci_c=(Ni)_2 \end{array}\right\}$$
...must conform to the magnetic loops and must be constant while each core is resetting



NOTES: 1. ASSUMED THAT $R_{wg}=0, R_s=R_{L1}, R_{L1} << R_{r1}, R_{r2}$

2. SCALE IS SUCH THAT CURRENT WAVEFORMS MAY ALSO BE INTERPRETED AS VOLTAGES $\frac{N_g}{N_c}i_cR_c, i_{r1}R_r, i_{r2}R_r$

Fig. 4. Theoretical waveforms, rectangular magnetic loop

A—Zero output B—Medium output
C—Almost full output D—Slightly beyond full output

---

* It would be best to characterize this circuit as "half-wave reversible a-c, reversible d-c output." There are actually two other types of reversible half-wave circuits: half-wave reversible a-c, nonreversible d-c output; and half-wave nonreversible a-c, reversible d-c output. The authors favor the shorthand notation RHWACDC, RHWAC, and RHWDC for these three types of circuits respectively. Fig. 3 illustrates these different possible output characteristics.

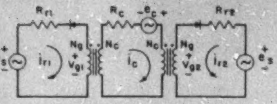


Fig. 5. Equivalent circuit during the negative or reset half-cycle

2. With control ampere-turns adding to reset ampere-turns in one core and subtracting in the other, one core will reach the shoulder of its loop and begin to reset before the other.

3. Operation is piecewise linear with three possible "pieces" or zones of linear operation from $-\pi$ to $-\pi/2(-\pi/2$ to 0 is symmetric to $-\pi$ to $-\pi/2$ unless saturation occurs).

These zones are:

a. both cores traveling along the top of the loops with all windings appearing as short circuits.

b. one core resetting with voltage across it while the other is still traveling along the top of its loop with its windings appearing as short circuits.

c. both cores resetting and absorbing voltage.

It is, of course, basic that the gate winding voltages V_{g1} and V_{g2} have zero average value or, expressed another way, that the reset and prefire voltage areas must be equal in steady-state operation. The flux in each core may be obtained by integrating V_{g1} and V_{g2}. One point of particular interest may be observed in the V_{g2} waveforms. During the positive half-cycle when both cores are unsaturated, nearly the full supply voltage appears across the gate windings of the two cores, and $V_{g1}=V_{g2}\doteq e_s$. When one core is saturated (core *1* in this case) and the other is not, the voltage appearing across the unsaturated gate winding is $V_{g2}=[(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg})]e_s$ assuming, as is done, that the inductance of the saturated gate windings drops to zero. It is apparent that V_{g2} can either continue smoothly or drop when core *1* saturates depending on the series resistance R_s and on whether $(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg})\leq 1$. In Fig. 4, it is assumed that $R_{wg}=0$ and $R_{L1}/(R_s+R_{L1})\doteq 0.5$. It is found that this ratio $(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg})$ has important consequences on triggering action.

Maximum output occurs when one core fires at 180 degrees and the other fires at 0 degrees. If one core reaches 180 degrees before the other reaches 0 degrees (or vice versa), the gain of the circuit is reduced at larger values of output as shown in Fig. 6(B). It should be noted that after one of the cores reaches a firing angle of 180 degrees, point *1* in Figs. 6(A) and (B), it proceeds to saturate on the reset half-cycle with the consequence that its dynamic operating loop changes from clinging-to-positive saturation to clinging-to-negative saturation as shown in Fig. 4(D). This change, which is quite similar to that occurring in the a-c doubler and other types of circuits, is accompanied by a time lag depending upon the ratio of the minor-loop height to 2 ψ_{sat} and to the magnitude of the input signal.[6] This delay, which may be made quite large, does not give rise to any discontinuity or double value in the static output versus input voltage characteristic and should not be confused with such discontinuity or triggering effect. On the basis of the square-loop model (or rectilinear model (B) in Fig. 1), ideal rectifiers, and resistive load, no triggering should occur. Winding and stray capacitance, leakage inductance, and other miscellaneous effects, which are unavoidable to some degree, in practice may give unpredictable results, but triggering does not occur in the ideal model.

First Observations of Triggering

Anyone who has worked with the half-wave bridge circuit or its equivalents has probably observed at least a slight amount of triggering when the amplifier output is reduced from oversaturation. This action is shown in Fig. 7(A). In one sense, however, the triggering is generally not a matter of practical importance to the equipment designer who, therefore, regretfully decides to forget about the phenomenon and go on to deal with more pressing problems. The fact is that operation in the saturated region must be avoided altogether if one is to preserve fast response.[2,4] It is often necessary to avoid oversaturation because of a decrease in output which may occur in this region as shown by dashed lines in Fig. 6. Limiting may be accomplished by limiting the input signal or by other more subtle means (limiting the amount of iron in the cores, for example). However accomplished, the limiting also elminates the triggering (generally). When the triggering is thus avoided, and as long as the triggering is rather minor to begin with, as it generally is, the designer is tempted to ignore the triggering, ascribing it to a number of vague "possibilities" affected by a host of nonideal conditions such as nonideal core leakage inductance, rectifier leakage and capacitance, finite nonlinear rectifier forward resistance, winding and stray capacitance, bridge rectifier backfiring during reset, etc. There is generally no shortage of possible causes of minor discrepancies in the operation of most magnetic-amplifier circuits, and the way in which positive feedback or inductive loading, intentional or unintentional, can cause instability is well described in the literature.[7-10]

Thus it was that triggering was observed, eliminated, and put out of mind on several occasions by the authors. Certain circuit changes were then made, however, which yielded triggering effects too enormous to ignore (see Fig. 7). Specifically it was found that an increase in the series resistor R_s or, more properly, in the ratio R_s/R_L increases the triggering proportionately and may increase it to the point where the circuit is simply bistable, i.e., yields full output of either one polarity or the other with no choice of output levels in between.

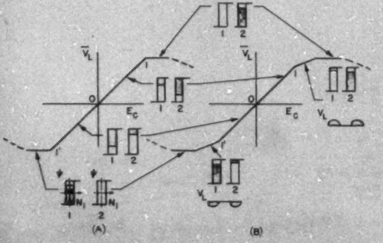


Fig. 6 (left). Voltage transfer curves

A—One core reaches 180-degree firing at same time the other reaches 0 degrees

B—One core reaches 180-degree firing before the other reaches 0 degrees

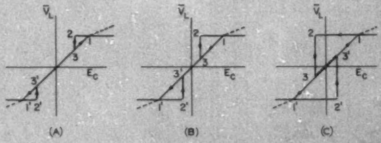


Fig. 7. Triggering

A—Slight
B—Moderate
C—Severe, through zero

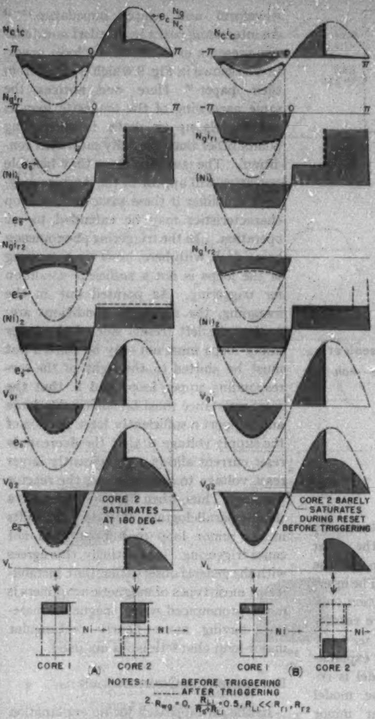

$e_c \frac{N_g}{N_c}$
$N_g i_c$
$-\pi$
0
π
$N_g i_{r1}$
e_s
$(Ni)_1$
$N_g i_{r2}$
$(Ni)_2$
V_{g1}
V_{g2}
V_L
CORE 2 SATURATES AT 180 DEG
CORE 2 BARELY SATURATES DURING RESET BEFORE TRIGGERING
Ni
CORE 1
CORE 2
(A)
(B)
NOTES: 1. ——— BEFORE TRIGGERING
------- AFTER TRIGGERING
2. $R_{wg}=0$, $\frac{R_{L1}}{R_s+R_{L1}}=0.5$, $R_{L1} << R_{r1}, R_{r2}$

Fig. 8(left). Theoretical waveforms showing triggering produced by magnetic loops of Fig. 2(C)

A—Output entering saturation

B—Output leaving saturation

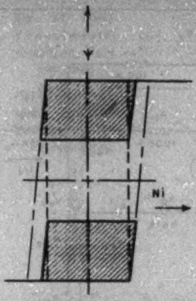


Fig. 9 (above). Magnetic core model from Fig. 8 of reference 16

Why one should wish to increase R_s is not an issue here, but it may be stated parenthetically that R_s may be used to limit circulating current in the load when both cores are saturated simultaneously and also that the amplifier gain may be increased as R_s is increased provided one is also free to modify and optimize turns and wire size on the gate windings. This latter statement is not evident without detailed analysis (to which there is no reference in the literature), but may be derived in a straightforward way none the less.

Triggering Explanation

Investigation has proved the reasons of triggering to be quite simple and to be entirely divorced from positive feedback or inductive loading. The causes are two fold. The first essential feature is that the minor hysteresis loops exhibit a shift similar to that shown in Fig. 2(C). This is confirmed beyond question by experimental data. The second essential feature is that the resistance in the reset circuit be large compared with the unsaturated impedance of the gate or reset winding. Under these conditions it is simple to show that triggering must occur, and experimental confirmation has been obtained. It can also be shown that triggering should and does occur if the control circuit is opened and output is controlled by varying one of the reset resistors. In fact, the underlying phenomenon causing triggering can be confirmed even further by observing its action in a single core.

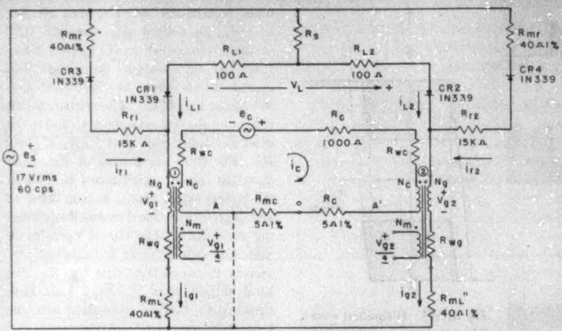
Theoretical Considerations

Waveforms based on an idealized core model, ideal rectifiers, and a value of $(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg}) \doteq 0.5$ are shown in Figs. 8(A) and (B). The core model used is shown in the lower part of figures. The minor magnetic loops remain rectangular but the ones clinging to $-\psi_{sat}$ are shifted approximately 10% of their total width in the $+(Ni)$ direction. At zero output the circuit will operate in the same way as shown previously in Fig. 4(A). The control circuit resistance is assumed to be matched to the reset resistance such that $R_c(N_g/N_c)^2=2R_r$. The solid waveforms of Fig. 8(A) show the case in which core *1* is firing at 90 degrees while core *2* is just firing at 180 degrees with the dynamic loop clinging to $+\psi_{sat}$.* The circuit operates in quite the same manner as explained previously in connection with Fig. 3. The total ampere-turns of core 2 have been sketched to show its correlation with the hysteresis loop.

Suppose that the control voltage is increased by an incremental amount. During the first reset half-cycle the voltage of V_{g2} is increased incrementally while the subsequent prefire voltage remains the same, being at its maximum value initially. If the core operated along the sides of a perfectly rectangular major loop, the operating path would merely circle slowly from $+\psi_{sat}$ to $-\psi_{sat}$ and make the transition shown in Figs. 4(C) and (D). Once at the bottom of the loop the core would barely saturate on the reset half-cycle so that reset and prefire areas again become equal and a new steady-state condition would be established without triggering. If the minor loop at $-\psi_{sat}$ is shifted to the right, however, the results are quite different. The total reset current required is reduced, the voltage drop across the reset resistance is decreased, and the peak reset voltage across the gate winding is increased as the operating path circles to the bottom of the loop. The prefire voltage does not change appreciably despite the increased prefire ampere-turns because of the low impedance of the load circuit, which means that $V_{g2} \doteq 0.5e_s$ even if the prefire current and the *IR* drops in R_s and R_{wg} are doubled or tripled. With the prefire voltage remaining approximately unchanged, the reset voltage area cannot increase in the steady state. Since the peak reset voltage has increased, there is no alternative but for the core to saturate at some angle appreciably before the end of the half-cycle in order to preserve the necessary equality of prefire and reset voltage areas. Note that the output remains unchanged at its maximum value and that triggering has occurred only in the reset voltage waveform.

---

* In general one would wish to have one core reach 0 degree at the same time that the other reaches 180 degrees, but the tendency to trigger is observed regardless.



NOTES: 1. R_{mL}' AND R_{mL}'' GIVE THE AMPLIFIER AN AFFECTIVE LOAD OF 280 Ω CT INSTEAD OF 200 Ω CT

2. CORES 1 AND 2 ORTHONOL MAGNETICS INC CAT. 50004-2A, 2 MIL 50-50 NICKLE-IRON GRAIN ORIENTED TAPE 1.250 IN OD 1.000 IN ID 0.250 IN H

N_c = 300 ±1% TURNS OF #36 WIRE, WINDING RESISTANCE 17 OHMS, R_{wc}

N_g = 2400 ±1% TURNS OF #31 WIRE, WINDING RESISTANCE 36 OHMS, R_{wg}

N_m = 600 ±1% TURNS OF #36 WIRE, WINDING RESISTANCE 33 OHMS

Fig. 10. Experimental circuit

If the control voltage is decreased now, output voltage triggering will occur, but not immediately. The reset voltage will gradually be decreased in peak amplitude while the core saturates later in the half-cycle and the reset voltage area remains constant. This continues until the reset saturation angle is pushed to 0 degrees. Throughout the process the output remains unchanged at its maximum level. On the output-input characteristic of Fig. 7(B), the circuit is traversing the path from point *1* to *2*. Point *2*, the point at which the control has been decreased until reset saturation occurs incrementally short of 0 degrees, is shown in the solid curves of Fig. 8(B). A further decrease in the control voltage sends the core operational path cycling up to $+\psi_{sat}$. If the core operated along the sides of a rectangular major loop, the process would be accompanied by a time delay but no triggering. For the minor loop shifted to the left, however, output triggering must occur. With the required reset current increased, the available reset peak voltage and voltage area across the gate winding is decreased. Since the prefire voltage area must equal the reset voltage area, in the steady state, the core must now saturate at some angle in advance of 180 degrees, and the output voltage makes a corresponding jump decrease out of saturation.

The sequence of events in the foregoing two paragraphs is not quite the whole story. Any triggering effect that begins to be exhibited in either the reset or prefire voltage of one core must necessarily alter the reset and prefire of the other core and be itself altered in turn because of the control circuit coupling. The interaction modifies the triggering action only slightly and the final results are readily sketched as shown.

Having gone thus far in the explanation of triggering, the core model is reexamined for a moment. The model used has perfectly rectangular minor loops. The one essential feature for triggering, however, is the shift in the left or reset side of the loop. It should be clear that this side of the loop may be vertical or sloping or may even change slope but triggering will occur so long as the average ampere-turns during reset show a decrease in going from upper to lower minor-loop operation. It might be thought that all necessary information is contained in the upper or lower loops alone, one set being a mere reflection of the other. This would be true except for the difference in exciting conditions during the reset and firing half-cycle. The effect of the different exciting conditions is brought out in the experimental data to follow, and may also be seen in minor loops of completely different character, as shown in references 10 and 11, despite the similarity of the core material and the major d-c and a-c magnetic loops. The width of any magnetic loop is very much a function of at least the first derivative of the flux (among other things not well understood) and is therefore very much a function of the exciting voltage waveform and source impedance.[13–15] An interesting paper by Batdorf and Johnson presents evidence of d-c loop operation as shown in Fig. 9 which is taken from their paper.[16] Here one notices the same narrowing of the magnetic loop as one travels up or down from starting points other than the fully saturated condition. The authors show that bistable operation will appear in the doubler magnetic amplifier if these same general loop characteristics may be extended to a-c operation. In the triggering phenomenon being dealt with here, however narrowing of the loops is not a sufficient condition for triggering. As pointed out in the foregoing, the necessary conditions are: 1. that the left (reset) side of the lower minor loops must not only be shifted but must be shifted to the right of the corresponding upper loop, and 2. that the reset resistance must be sufficiently large and support a sufficiently large portion of the supply voltage so that the decrease in reset current allows a significantly larger reset voltage to appear across the reactor winding. Thus, when the major loop has a sloping parallelogram characteristic, shifts in the minor loop may occur and not cause triggering. Interestingly, this agrees with the general observation that instability in most types of magnetic amplifiers is more pronounced when magnetic materials having more nearly rectangular major-loop characteristics are used.

EXPERIMENTAL CONFIRMATION

In the initial search for an explanation of the triggering phenomenon, which seemed to persist despite gradual elimination of other postulated causes, the authors experimented on paper with the effects of different core models and arrived at the reverse-slope effect, just reviewed, as a possible cause of triggering.

Experimental evidence was then sought to confirm or reject this possible explanation. Experimental corroboration is not immediately evident since one cannot ordinarily view the total core ampere-turns waveforms $(Ni)_1$ and $(Ni)_2$ of Figs. 8(A) and (B). The experimental circuit of Fig. 10 was soon arranged wherein current-measuring resistors proportional to the winding turns were placed in the control and load circuits. If one uses the resistor values shown in Fig. 10 and connects the control and output circuits along the dashed line, the voltage *OB* is proportional to the total ampere-turns around core *1*. In a similar fashion (but using a different control-load circuit connection) the total ampere-turns on core *2* may be obtained by a single voltage waveform.

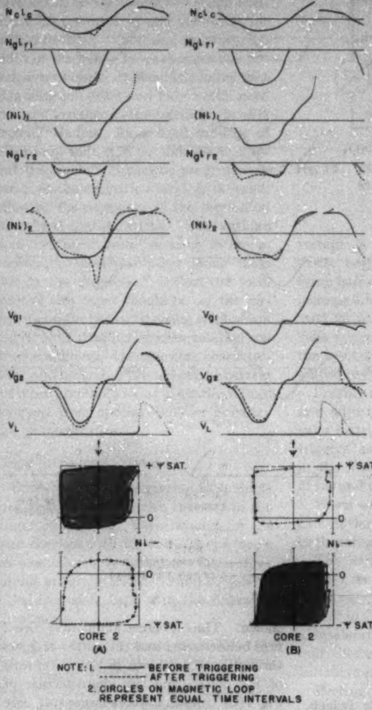


Fig. 11 (left). Experimental waveforms, R_s = 100 ohms

A—Output entering saturation
B—Output leaving saturation

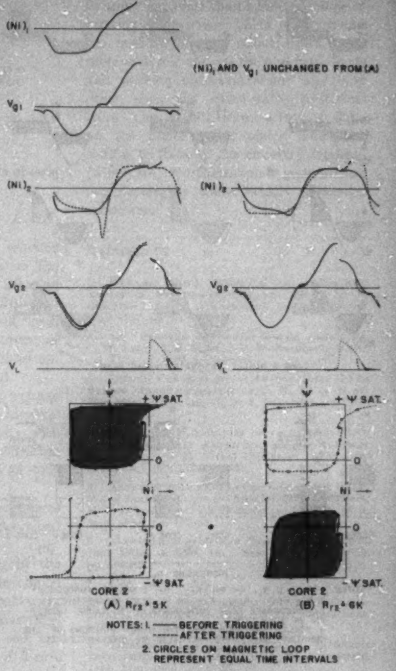


Fig. 12 (right). Experimental waveforms, R_s = 100 ohms. Control circuit open; output controlled by varying R_{r2}

A—Output entering saturation
B—Output leaving saturation

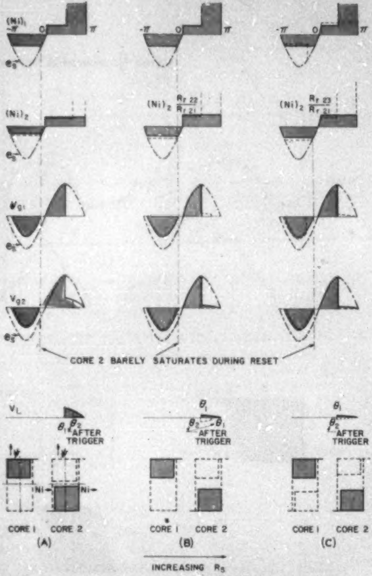
Of course, the individual winding currents may be obtained and added graphically, but the single voltage trace on an oscilloscope gives a convenient, immediate continuous monitor of the dynamic loop width while making exploratory circuit parameter changes. Pickoff windings were used to obtain coil voltages unaffected by *IR* drops. Fig. 11 presents waveforms, taken from oscilloscope photographs, together with magnetic-loop paths derived from the current and voltage (graphically integrated) waveforms. These waveforms should be checked against those of the theoretical model shown in Fig. 8. The data confirm the reverse-slope effect quite nicely. The jog in the rising portion of the magnetic loop is of minor importance. It results from the sudden drop in voltage across the load winding of core *2* as core *1* saturates. As previously discussed, the general shape of the loop during the firing or positive half-cycle has negligible effect on circuit operation. The jog does serve to show that the upper and lower loops need not be symmetric to one another.

Output Controlled by Varying Reset Resistance

As further confirmation of the cause of triggering it was decided to leave the control circuit open and look for the triggering effect with the output controlled by varying one of the reset resistors. Reset voltage across a gate winding can be effectually varied in this manner to control the output voltage of the amplifier. This method of operation has two advantages. First, it eliminates the possibility of control circuit interaction as a cause of triggering. Second, it allows the total coil current to be observed across a single current-measuring resistor and eliminates the possibility of incorrect scaling of ampere-turns contributions of the control circuit to the total ampere-turns because of an erroneous turns count, etc. Output triggering is predicted much as in the previous case except that the output is now a function of one of the reset resistances rather than control voltage. The waveforms of Fig. 12 exhibit the triggering action clearly and again confirm the reverse-slop effect in the magnetic loop. The fact that these waveforms were obtained with the control circuit open clearly proves that the control circuit interaction plays no essential role in the triggering effect. With the output going into saturation, triggering tends to occur somewhat before the firing angle of core *2* reaches 180 degrees as shown in Fig. 12(A) causing a noticeable discontinuity in the output.

Effect of Large Series Resistance R_g

The case dealt with thus far is $(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg})\doteq 0.5-0.6$. After the first core saturates during the positive half-cycle, the voltage across the load winding of the second drops from e_s to approximately $(0.5-0.6)e_s$. The effect of varying R_g is examined now. Experimentally it is found that large values of R_s or R_s/R_{L1} aggravate the triggering action and can even cause the output to jump through zero as shown in Fig. 7(C).



NOTES: 1. $R_{wg} = 0$, $R_{L1} << R_{r1}, R_{r2}$. R_{r1} CONSTANT THROUGHOUT.
2. $R_{r2} = R_{r21}, R_{r22}, R_{r23}$ IN (A) (B) (C) RESPECTIVELY. R_{r2} INCREASED ONLY INCREMENTALLY TO CAUSE TRIGGERING IN EACH CASE.
3. SCALES ARE SUCH THAT THE RESET PORTIONS OF THE CURRENT WAVEFORMS MAY ALSO BE INTERPRETED AS $i_{r1} R_{r1}$ AND $i_{r2} R_{r2}$.

Fig. 13 (left). Theoretical waveforms. Triggering with increased series resistance. Control circuit open; output controlled by varying R_{r2}

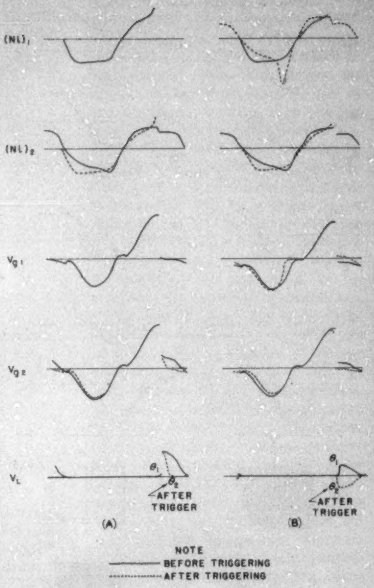


Fig. 14 (right). Experimental waveforms. Triggering with increased series resistance. Control circuit open; output controlled by varying R_{r2}
A—R_s = 500 ohms
B—R_s = 1,200 ohms

The reasons are not difficult to see. In Fig. 13 waveforms are drawn to show the effect of increased R_S on the triggering which occurs in decreasing the output from saturation. In this case the control circuit is open with the output controlled by varing reset resistance R_{r2}. At zero output any change in the series resistance has no effect on the circuit other than to reduce the current which flows after both cores are saturated: assuming, of course, that magnetizing current through R_S gives negligible voltage drop. As the output is varied, whether by supplying a control voltage or by varying one of the reset resistors, the effect of the series resistor becomes evident. R_S can have no effect on the reset process in either case since the load circuit is kept open-circuited by its rectifiers during the reset half-cycle. On the positive half-cycle R_s makes itself felt after the first core saturates (core *1* at θ_1, in this case). At this time the voltage across the other core drops to $(R_{L1}+R_{wg})/(R_s+R_{L1}+R_{wg})e_s$ as discussed in previous sections. Since the prefire voltage area of this second core must equal the reset voltage area of the preceding half-cycle (which has been unaffected by R_s), it follows that the lower the voltage after θ_1, the later in the cycle will the core saturate. It also means that the greater will be the angular jump in V_{g2} (and the greater the change in output pulse width) for any abrupt change in reset during triggering. It should be noted that the presence of R_s decreases the magnitude of the output pulse and that in terms of output pulse area the triggering is not necessarily greater.

As long as the firing angle θ_2 of the second core does not attempt to trigger to less than θ_1, V_{g1} remains unchanged. If R_s is made larger, the unique circumstance shown in Fig. 13(B) is encountered. Here the prefire voltage cannot decrease sufficiently unless it saturates before the other core fires at θ_1. It does just this. Now, however, V_{g1} has been tampered with since its value must drop after its companion saturates. Since its reset has been altered little or not at all, the prefire voltage area must stay the same. The only way in which this can happen is for core *1* to saturate later in the cycle as shown. Thus it comes about that core *2* fires before core *1* and the output triggers through zero as shown in the waveforms of Fig. 13(B) and the characteristic of Fig. 7(C). As R_s is made larger still, core *1* will saturate later in the cycle until it reaches 180 degrees. Finally the reset voltage of core *1* exceeds the prefire voltage even though the core does not fire until 180 degrees. Core *1* must then travel down the loop and saturate during reset. This case is shown in Fig. 13(C).

Experimental waveforms taken from oscilloscope photographs are shown in Fig. 14. These waveforms agree well with the theoretical ones of Figs. 13(A) and (C).

EFFECT OF VARYING EXCITATION LEVEL

One would expect that the height of the the minor loops in relation to the major loops might have a great bearing on the magnitude of the sideways shift of the minor loops and on the triggering effects just described. Specifically, one might expect the shift to decrease or disappear as the size of the minor loop at maximum output approaches that of the major loop, the condition generally described as "normal excitation." Laboratory tests indicate that the triggering does decrease although in many instances triggering is observed at or very close to normal excitation as reported in the following for single-core tests. The way in which the width of the minor loops changes has not been determined. It should be noted that triggering intensity and loop width need not vary together as the excitation level is altered. Indeed, for a fixed amount of minor-loop shift, it is not difficult to show that triggering will become progressively worse as the excitation level is reduced, following the reasoning of the section on theoretical considerations. The authors have refrained from making intensive loop-width experimentation along these lines in the conviction: 1. that the focal point of this paper should be on the circuitry and the basic triggering mechanism and 2. that detailed studies relating exciting conditions, dynamic core characteristics, and basic core physics can be carried out more profitably along generalized lines divorced from special amplifier circuitry and characteristics.

Tests on Single Cores

Tests to observe triggering in a single core were conducted with the core in the circuit of Fig. 15. Three wound toroids were checked with 60- and 400-cps (cycle per second) supply voltages. These three toroids are grain-oriented 2-mil 50%-50% nickel-iron-alloy tape with the following dimensions:

Core *1*...1-inch ID (inside diameter), $1^1/_4$-inch OD (outside diameter), 1/4-inch height, ID/OD ratio = 0.8
Core *2*...$1^1/_2$-inch ID, 2-inch OD, 1-inch height, ID/OD ratio = 0.75
Core *3*...$1^1/_4$-inch ID, 1.875-inch OD, 1/2-inch height, ID/OD ratio = 0.668

With the cores set up as in Fig. 15, 60-cps normal excitation swings the cores from positive to negative saturation. Under this condition the core does not enter the saturated region, and the usual square wave of exciting current is observed across R_2. Next the reset supply voltage E_r is decreased by 3%. A small trigger is observed in the current waveform during the positive or firing half-cycle. The trigger consists of a peak (at the end of the firing half-cycle) approximately 133% higher than the exciting level and 18 degrees in duration. The supply voltage is then reduced 30% below the normal excitation level. The current waveform across R_2 shows saturation on the reset half-cycle. As the reset voltage is reduced, a point is reached where triggering into saturation on the firing half-cycle takes place with duration increased to 36 degrees and amplitude at least 20 times the exciting current level. This trigger, with careful adjustment of the reset voltage magnitude, can be observed to take as long as 1 second.

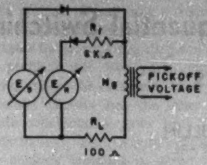


Fig. 15. Experimental circuit used to observe triggering in a single core

Further reduction of the supply voltage and adjustment of the reset voltage results in triggers of greater severity with triggers as large as 90-degree duration in the firing half-cycle recorded at 40% of normal excitation.

When excited at 400 cps, these cores exhibit the same phenomenon as described in the foregoing.

The explanation of triggering in the single core is the same as that developed previously for the reversible half-wave circuit in the section entitled "Theoretical Considerations." In the single-core circuit, however, any possibility of 2-core interaction is eliminated as a cause of triggering. The tests at various levels of excitation show that the minor-loop shift is not a phenomenon restricted to reduced excitation but may occur for minor loops which are very nearly as large as the major loop itself.

Conclusions

A new mode of magnetic-amplifier instability has been described and attributed to a reverse-slope effect in the locus of the minor magnetic loops of 50%-50% nickel-iron rectangular-loop core material. Theory and supporting experimental data have been presented to show that this unusual loop effect is actually present and is responsible for triggering of the amplifier as its output voltage enters and leaves its saturated maximum level. The reader is again cautioned that no model of minor- or major-loop operation can be expected to apply for other than a limited range of exciting conditions. There is, or should be, nothing strange about bows, jags, reverse slopes, etc., in the magnetic loops and nothing universal about any loop other than the quasi-static hysteresis loop. One should expect a reverse slope on the minor loops under conditions similar to those of the circuitry reported (which are nevertheless quite general) and not necessarily under widely dissimilar conditions.

References

1. DYNAMIC CORE BEHAVIOR AND MAGNETIC AMPLIFIER PERFORMANCE, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 229–40.

2. AN IMPROVED MAGNETIC SERVO AMPLIFIER, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *Ibid.*, vol. 71, pt. I, 1952, pp. 281–89.

3. DESIGN CONSIDERATIONS OF THE HALF-WAVE BRIDGE MAGNETIC AMPLIFIER, C. W. Lufcy, H. H. Woodson. *Ibid.*, vol. 73, pt. I, July 1954, pp. 220–26.

4. PRELIMINARY ANALYSIS OF THE HALF-WAVE BRIDGE MAGNETIC AMPLIFIER, H. H. Woodson. *Report 3596*, U. S. Naval Ordnance Laboratory, White Oak, Md., Feb. 1954.

5. ALTERNATING CURRENT CONTROL OF THE HALF-WAVE BRIDGE MAGNETIC AMPLIFIER, T. Bernstein, N. L. Schmitz. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 433–38.

6. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 150–52.

7. *Ibid.*, pp. 199–208.

8. MAGNETIC AMPLIFIERS IN BISTABLE OPERATION, L. A. Finzi, G. C. Feth. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 592–98.

9. INDUCTIVE LOAD INSTABILITY IN MAGNETIC AMPLIFIERS, H. I. Leon, Allen B. Rosenstein. *Ibid.*, Mar. 1955, pp. 8–20.

10. MAGNETIC AMPLIFIERS WITH INDUCTIVE D-C LOAD, H. F. Storm. *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 467–75.

11. *Technical Bulletin TC-101A*. Arnold Engineering Company, Marengo, Ill., 1953, pp. 12–13.

12. MAGNETIC AMPLIFIER ANALYSIS USING A GENERALIZED MODEL FOR THE SATURABLE REACTOR CORE, H. H. Woodson. *WADC Technical Note 56-198*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

13. DESCRIBING FUNCTIONS FOR THE RESET OF MAGNETIC AMPLIFIER CORES, G. C. Feth. *Technical Report 24*, Magnetic-Amplifiers, Carnegie Institute of Technology, Pittsburgh, Pa.

14. MAGNETIC CHARACTERISTICS PERTINENT TO THE OPERATION OF CORES IN SELF-SATURATING MAGNETIC AMPLIFIERS, R. W. Roberts. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

15. LOW-IMPEDANCE OPERATIONAL CHARACTERISTICS OF TOROIDAL CORES USED IN MAGNETIC AMPLIFIERS, C. L. Boyajian. *Ibid.*, vol. 75, pt. I, July 1956, pp. 307–13.

16. AN INSTABILITY OF SELF-SATURATING MAGNETIC AMPLIFIERS USING RECTANGULAR LOOP CORE MATERIALS, S. B. Batdorf, W. N. Johnson. *Ibid.*, vol. 72, pt. I, 1953, pp. 223–28.

17. *Catalog TWC-100*. Magnetics Inc., Butler, Pa., p. 14.

Reliable Stochastic Sequential Switching Circuits

ALBERT A. MULLIN
ASSOCIATE MEMBER AIEE

Synopsis: A procedure based upon probabilistic logics is developed for the analysis and synthesis of reliable asynchronous relay sequential switching circuits, subject to a set of statistical postulates. The reliability of a sequential circuit is measured as proportional to a logarithm of the probability of error for certain state transitions of the circuit involved. The "ces" as a unit of reliability is proposed by the author. An example of the analysis and synthesis procedure is given.

THE PURPOSE of this research is to investigate certain aspects of the nature and effects upon circuit action of a probabilistically specified malfunction attributed to the contacts used in the realization of relay asynchronous sequential circuits.

Relevant Background

At present, the principal published papers on the use of probabilistic logics as a mathematical tool for effecting the synthesis of reliable logical systems are quite recent, and two in number: the first by Von Neumann[1] and the second by Moore and Shannon.[2] Certain parts of Von Neumann's paper are a ramification and extension of the concepts contained in a paper by McCulloch and Pitts.[3] Von Neumann treats logical systems composed wholly of one or the other of two specific components. The treatment is carried out in quite a general mathematical fashion. A lower bound on the error level for each of the components of $\epsilon<1/6$ is needed in Von Neumann's model to insure receiving a more reliable automatum by the proper M-truplication of its components.

Moore and Shannon explicitly consider relays as components used to construct certain primarily combinational, logical systems.[2] In their model, each relay comprising the circuit is looked upon as a noisy binary information channel, which is not necessarily symmetric. Relays as components are less general than the components treated by Von Neumann. Nevertheless, many useful results still appear, since Moore and Shannon have shown that "by using a sufficiently large number of relays in the proper manner ... circuits can be built that are arbitrarily reliable, regardless of how unreliable the original relays are."

Terminology

Because of their relative simplicity, relays will be used as a vehicle for demonstrating the theory. However, it is intended that much of the theory itself be independent of specific components. The word "classical" will be used in reference to a set of deterministic (nonprobabilistic) procedures for synthesizing sequential circuits.

In two significant papers, Huffman has developed a deterministic model for synthesizing classical sequential circuits[4] and a procedure for synthesizing hazard-free classical combinational circuits. These two fundamental papers will be referred to freely by the author.[5] (An engineering textbook covering these subjects among many other topics in classical switching circuits is reference 6.)

The concept of primary and secondary relays will be carried over from the classical theory of relay sequential circuits. Thus, a primary relay in the probabilistic theory is one whose relay coil is connected directly to an input line of the circuit and a secondary relay is any other relay that is utilized in the circuit.

In the following, the author will be concerned with the probabilistic model unless a specific statement to the contrary appears.

Normally open and normally closed contacts will be defined by means of a binary information channel.[7] Such a binary channel is shown in Fig. 1, where, for example, A is the probability that a specified contact on a specified relay is open, on the hypothesis that the relay coil is de-energized.

Definitions

1. A relay contact will be called normally open if and only if $A-B>0$.
2. A relay contact will be called normally closed if and only if $A-B<0$.
3. A relay contact will be called random if and only if $A-B=0$.

From the probabilistic point of view, it may be said that normally open and normally closed contacts differ only in degree rather than in kind.

4. The conditional probability of the event A, given that the event B has occurred, denoted by $P(A|B)$, is defined as the joint probability of the events normalized to the probability of event B.

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

5. Several events are defined to be statistically independent if the probability of occurrence of any one of them is not affected by the supplemental knowledge concerning the occurrence of any number of the remaining events. For two events A and B this may be expressed in the form $P(A|B)=P(A)$, or $P(B|A)=P(B)$. A more concise definition of statistical independence appears in reference 8.
6. The state of a sequential circuit is defined by the ordered set of 2-state variables $X_1, X_2, \ldots X_M, Y_1, Y_2 \ldots Y_N$ where X_1 is an ordered set of 2-state response variables $x_{11}, x_{12}, \ldots x_{1Q}$ of the Q contacts utilized from the primary relay $x_1, \ldots$, etc., and Y_N is an ordered set of 2-state response $Y_{N1}, Y_{N2}, \ldots Y_{NR}$ of the R contacts utilized from the secondary relay Y_N. (The significance of this abstract definition appears in the section "Analysis.")
7. A transition probability is defined as the conditional probability of the circuit going to a given state on the hypothesis that it was in a given state. As an example, $P(S_2|S_1)$ is the transition prob-

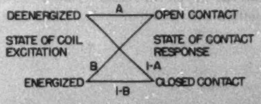


Fig. 1. Noisy binary channel for defining normally open and normally closed contacts

---

Paper 58-854, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted December 12, 1957; made available for printing April 25, 1958.

ALBERT A. MULLIN is with the University of Illinois, Urbana, Ill.

This paper is a condensed version of a master's thesis entitled "Probabilistic Logics in the Synthesis of Reliable Sequential Circuits," which was submitted on July 12, 1957, to the Electrical Engineering Department of the Massachusetts Institute of Technology, Cambridge, Mass. The author acknowledges the constructive criticism of Prof. C. E. Shannon, who supervised the thesis. However, the author accepts full responsibility for any errors. He is indebted, for the "feedback" received, to the switching circuits group of the Digital Computer Laboratory at the University of Illinois, through a series of three lectures on the subject of the paper. The author dedicates this paper to his mother and father.

ability of the circuit going to state 3, given that the circuit was in state 1.

8. A stochastic matrix is defined as a square matrix whose elements are transition probabilities. The element P_{JK} of the stochastic matrix is at the intersection of row J and column K and is identical with $P(S_K|S_J)$.

Postulates

1. Only one input line may change state at any given time and the same or another input line will not change state until the circuit settles down.

2. Each relay has only one relay coil, and will be assumed to have a defining probability for its normally open contacts so that $A \geq 0.9$ and $B \leq 0.1$. The defining probability for the normally closed contacts will be such that $A \leq 0.1$ and $B \geq 0.9$.

This assumption should not detract very seriously from the general theory, and certainly not from applications utilizing commercially available relays. However, random contacts and excessively crummy[3] (low quality) relays will be excluded from interest.

3. Each utilized contact of every relay functions as if it were statistically independent of the remaining contacts of the relay concerned.

4. When the state of excitation to a relay coil is changed, each of the contacts on the relay change their response state, once at most, according to the statistical hypothesis. Thus, malfunction due to contact bounce will not be considered.

5. On a relay whose coil has had its state of excitation changed, the probability that any two contacts change their response state simultaneously is zero.

6. The members of the set of permutations of sequence in which the utilized contacts on each relay function or malfunction are equiprobable.

Analysis

The probabilistic-logical analysis problem involves being given a specific sequential circuit which is assumed to be composed of error-prone relays whose contacts function in accordance with the foregoing postulates. The statistical error associated with the circuit action is assumed to be localized with the component relays and to be due to the M-truplication of the variables. The composition of variables introduces no noise in this case. The relay as a component should be contrasted with potential-level devices such as transistors and cryotrons, where just the opposite is true.

The problem of analyzing a sequential circuit by probabilistic-logical techniques is effected by decomposing the given analysis problem into two parts, one involving only the classical theory of sequential circuits and the other only certain manipulation laws from probability theory. Each of these less formidable analysis problems is solved in its own environment and their individual results are fused together to give the solution to the original analysis problem. Thus, the probabilistic-logical analysis procedures will be a natural generalization of the classical analysis procedures.

The following is an analysis procedure for circuits free of static hazards. (The effects of static hazards upon circuit action is deferred for the present. However, an individual treatment of the quantitative effects of static hazards upon circuit action will be given later in the paper.)

Step 1. For the present discussion, the probabilistic circuit should have no kinds of hazard present. From the given sequential circuit obtain the logic properties of the corresponding classical circuit. The corresponding classical circuit is obtained, from the probabilistic circuit by using the defining probability for each utilized contact, to ascertain whether the contact is normally open or normally closed. Then, the contact so defined is considered normally open or normally closed in the classical sense. The method developed by Huffman[4] may be used to obtain the work statement of the corresponding classical circuit. Let the circuit contain M primary relays and N secondary relays. Then, the rectangular array formed by heading the columns with all possible states for an ordered set of primary response variables, and the rows with all possible states for an ordered set of secondary response variables, is a $2^N \times 2^M$ matrix. This flow matrix[4] is sufficient to describe the circuit action for the corresponding classical circuit.

Step 2. Consider the probabilistic portion of the analysis. Suppose that the state of excitation to any single relay coil changes. The response state of each normally open or normally closed contact is no longer tightly coupled to the excitation state of its relay coil. Each contact on the relay whose coil excitation changed state may be open or closed independently of the remaining contacts.

Let the primary relay response variable x_1 be β_1-truplicated, x_2 β_2-truplicated, $\ldots$, x_M β_M-truplicated, and the secondary relay response variable y_1 α_1-truplicated, $\ldots$, y_N α_N-truplicated. The state of the probabilistic sequential circuit is defined by the ordered set

$$x_{11}, x_{12} \ldots, x_{1\beta_1}, \ldots, y_{N1}, y_{N2} \ldots y_{N\alpha_N}$$

when the specific state (0 or 1) of each of the response variables is inserted. A double subscript notation is used with each of the 2-state response variables. The first subscript refers to the correspondingly numbered relay coil involved, and the second to the particular utilized contact controlled by the state of excitation applied to that coil.

Hence, if the general probabilistic problem was described by a rectangular array constructed in a manner similar to the one used for the classical problem, a

$$2^{\sum_{j=1}^{N} \alpha_j} \times 2^{\sum_{j=1}^{M} \beta_j}$$

matrix would be sufficient to describe the circuit action. This fundamental matrix will be called the primal matrix. It contains P elements, where

$$P = 2^{\left(\sum_{j=1}^{N} \alpha_j + \sum_{j=1}^{M} \beta_j\right)}$$

Hence, the numbr of elements in the stochastic matrix is S where $S = P^2$. Thus, even for relatively simple problems, it is not feasible to solve for the complete stochastic matrix without the aid of a high-speed data-processing machine.

Consider the less ambitious problem of a very special stochastic matrix. This stochastic matrix will involve only the transitions between those states, as represented in the primal matrix, that agree with the stable states of the corresponding classical circuit. Actually, these transitions are the more important ones for many practical problems, because the corresponding classical circuit is a convenient standard for reliability.

Step 3. Proceed to find the elements of the special stochastic matrix, hereafter called "the stochastic matrix." Such matrices are computable from the defining probabilities for the contacts, the geometry of the circuit, and the set of operational postulates. The computation is done as follows:

a. Inspect the flow matrix of the corresponding classical circuit. Find all of the transitions from each of the stable states to any other stable state that can be made by the change of a single input variable. These will be called allowable transitions.

b. Inspect the primal matrix. Find all of the transition probabilities between each of the states in the primal matrix corresponding to the allowable transitions between stable states in the flow matrix.

Table I. Probability Distribution for a Sequence Series

	x_{21}	x_{22}		x_{12}	x_{11}
$\frac{(0.1)^2}{2}$	0	0	$\frac{(0.1)^2}{2}$	0	0
$\frac{(0.1)(0.9)}{2}$	0	1	$\frac{(0.1)(0.9)}{2}$	0	1
$\frac{(0.9)(0.9)}{2}$	1	1	$\frac{(0.9)(0.9)}{2}$	1	1
$\frac{(0.9)(0.1)}{2}$	1	0	$\frac{(0.9)(0.1)}{2}$	1	0

c. To become more specific, consider the typical allowable transition from stable state α to stable state β as would be represented in the flow matrix. Let state γ and state δ in the primal matrix correspond respectively to the stable state α and the stable state β in the flow matrix. Each of the response variables, from the primary relay whose excitation is changed most recently, is statistically independent of the remaining response variables.

d. Find the column λ of the primal matrix headed with the particular set of input response states corresponding to state δ.

e. The probability μ of the circuit going to a state which is in column λ of the primal matrix may be found by applying definition 5 to the response state of the contacts concerned. This involves only contacts from the primary relay whose coil excitation has been changed. Contacts from any other relay or relays do not change their response state, at least, not yet.

f. Consider the states in column λ of the primal matrix. Find the probability ν of going to state δ. The probability ν is specifically conditioned on the state being in the column λ of the primal matrix. The probability ν is found by determining the new condition of excitation to each of the secondary relay coils. If state γ and state δ are in the same row of the primal matrix, the secondary coil excitation will not change. Apply definition 5 for the new condition of secondary coil excitations as regards the secondary response variables corresponding to state δ.

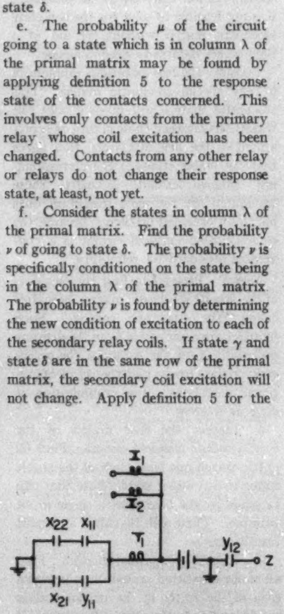


Fig. 2. Sequential circuit to be analyzed

g. The transition probability $P(\delta|\gamma)=\mu\times\nu$ is entered into the stochastic matrix at the intersection of row γ and column δ.

The transitional probabilities obtained by this procedure are lower bounds to the probabilities of receiving the output state identical to that of the corresponding classical circuit. Any transition probability corresponding to a nonallowable transition between stable states of the flow matrix will be undefined. A dash will be inserted into the stochastic matrix for this transition probability.

Let $[\text{ж}_{JK}]$ denote the stochastic matrix and $[\text{Э}_{JK}]$ denote the error matrix. Then, symbolically

$$(\text{э}_{JK})\equiv(1-\text{ж}_{JK})$$

$$\begin{bmatrix} \text{э}_{11} & \text{э}_{12} & \ldots & \text{э}_{1A} \\ \text{э}_{21} & \text{э}_{22} & \ldots & \text{э}_{2A} \\ \vdots & \vdots & \ldots & \vdots \\ \text{э}_{A1} & \text{э}_{A2} & \ldots & \text{э}_{AA} \end{bmatrix} \equiv \begin{bmatrix} (1-\text{ж}_{11}) & (1-\text{ж}_{12}) & \ldots & (1-\text{ж}_{1A}) \\ (1-\text{ж}_{21}) & (1-\text{ж}_{22}) & \ldots & (1-\text{ж}_{2A}) \\ \vdots & \vdots & \vdots & \vdots \\ (1-\text{ж}_{A1}) & (1-\text{ж}_{A2}) & \ldots & (1-\text{ж}_{AA}) \end{bmatrix} \quad (1)$$

Of course, if any particular ж_{JK} is undefined, then $\text{З}_{JK}=1-\text{ж}_{JK}$ is likewise undefined.

The measure of reliability will be taken as proportional to a logarithmic function of the error probabilities, which is used because: 1. it has intuitive significance (i.e., zero probability of error gives a "large" reliability and unity probability of error gives zero reliability); 2. it is more suggestive of the order of magnitude of the equipment needed to achieve a given reliability; and 3. it is indicative of the entropy of the error involved.

Let (P_{JK}) denote the reliability matrix. Let the mark "o" denote a binary composition law with the following property:

$$[P_{JK}]\equiv -\log_{o}[\text{э}_{JK}]$$

$$\begin{bmatrix} P_{11} & \ldots & P_{1A} \\ \vdots & \ldots & \vdots \\ P_{A1} & \ldots & P_{AA} \end{bmatrix} \equiv -\log_{o} \begin{bmatrix} \text{э}_{11} & \ldots & \text{э}_{1A} \\ \vdots & \ldots & \vdots \\ \text{э}_{A1} & \ldots & \text{э}_{AA} \end{bmatrix} \equiv \begin{bmatrix} -\log \text{э}_{11} & \ldots & -\log \text{э}_{1A} \\ \vdots & \ldots & \vdots \\ -\log \text{э}_{A1} & \ldots & -\log \text{э}_{AA} \end{bmatrix} \quad (2)$$

Since the stochastic matrix gave lower bounds to the probabilities of obtaining an output state agreeing with the corresponding classical circuit, the reliability matrix is a lower bound for the reliability with which this circuit action is obtained.

A specific base to the logarithmic measure of reliability has not yet been made. The author proposes using logarithms to the base 2 and calling the unit of reliability the "ces," in honor of Prof. C. E. Shannon, who has carried out considerable fundamental research in the field of reliable communication.

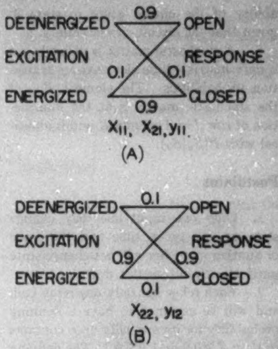


Fig. 3. Binary symmetric channel used for defining contacts shown in circuit of Fig. 2

A—Normally open contacts
B—Normally closed contacts

If the elements of the reliability matrix are not too much dispersed, a useful measure of reliability for circuit action is the single number $P_0=\min P_{JK}$.

Static Hazards

Let a sequential circuit utilize K contacts from a relay A_1. By the "sequence-series on the relay A_1" is meant the set comprised of the $(K!)$ permutations of the ordered set of utilized contacts A_{11}, A_{12}, …, A_{1K} where each of the permutations contains 2^K members. Each permutation corresponds to the time sequence in which the K utilized contacts from relay A_1, either function in accordance with the classical circuit or malfunction. The 2^K members of each permutation is the number of ways in which the K utilized contacts either function or

Table II. Excitation Matrix (Y_1) and Flow Matrix (F) for Corresponding Classical Circuit Shown in Fig. 2

		X_1X_2			
		00	01	11	10
Y_1	0	0	0	0	1
	1	0	1	1	1

$$Y_1=X_1X_2'+X_2Y_1$$

	X_1X_2			
	00	01	11	10
Y_1	①	③	⑤	2
	1	⑥	④	②

F

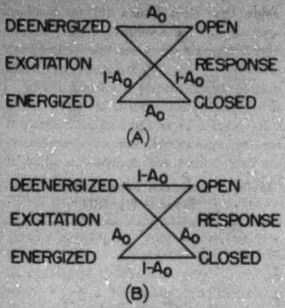


Fig. 4. Binary symmetric channel used for defining contacts to be utilized in synthesis of circuits

A—Normally open contacts
B—Normally closed contacts

malfunction. The probability distribution over the ensemble of $2^X \times K!$ sequence-series may be found by applying postulates 3 and 6. Table I shows an ensemble of sequence series for $K=2$, for a normally open contact with a defining probability of $A=0.9$ and $B=0.1$.

The actual procedure involved in analyzing the effects of static hazards upon the reliability of circuit action is as follows:

1. Take the given sequential circuit and find the corresponding classical circuit.

2. Inspect the excitation (Y) matrix[4] for the control of the secondary relay coils of this classical circuit for any static hazards. This may be done by a method developed by Huffman.[5]

3. Transitions between states of the primal matrix, which correspond to allowable transitions in the flow matrix for which a static hazard can be present, must be re-evaluated. By using the defining probabilities for the utilized contacts on the primary relay, the probability distribution over the ensemble of sequence series is found. Let φ be the sum of the probabilities of those members of the sequence-series which correspond tò the condition of no hazards nor malfunction. Then, 3 where $3 = 1-\varphi$ is the probability of error due to the effects of static hazards and M-truplication of variables for the particular sequence involved. This error probability is substituted into the error matrix.

4. If, after treating the primary relay whose change of excitation involves a static hazard, any secondary response of contacts involves a static hazard, this is treated by finding the probability distribution over the ensemble of sequence series for the utilized contacts on the secondary relay involved.

Table III. Reliability Matrix (P) for Circuit Shown in Fig. 2

	1	2	3	4	5	6
1	..	1.88	2.30	..	..	..
2	1.88	..	..	0.75	..	..
3	2.39	..	..	..	3.32	..
4	..	0.75	..	..	..	3.32
5	..	1.54	3.32	..	..	..
6	1.54	..	..	3.32	..	..

P

Let A equal the probability of error due to static hazards and the M-truplication of variables in the primary relay, and let B equal the probability of error due to static hazards and the M-truplication of variables in the secondary relay. The probability B is conditioned on the effects of the primary relay. Then, the probability C of at least one error for the state transition described is $C=A+B-A\times B$.

For this more general condition, C is substituted into the error matrix. If the corresponding classical circuit is free of static because all the prime implicant tie sets and cut sets have been added,[8] the probabilistic circuit can still have error due to the effects of static hazards. This follows because the added prime implicant tie sets and cut sets are subject to malfunction in the probabilistic circuit in accordance with the statistical postulates.

It can be shown fundamentally that an added prime implicant tie set increases the probability of having at least one tie set through the network. However, it decreases the probability of having at least one cut set. Thus, an added prime implicant tie set can increase the probability of error, due to the M-truplication of variables, for certain state transitions. The added prime implicant tie set will decrease the probability of error due to the effects of static hazards.

EXAMPLE OF ANALYSIS PROCEDURE

Consider the sequential circuit shown in Fig. 2. The defining probabilities for each of the contacts are given in Fig. 3(A) and (B). An excitation and flow matrix for the corresponding classical circuit is given in Table II. The word statement of the logical properties of the corresponding classical circuit shown in Fig. 2 is, on a transmission basis, the following: $Z=0$ if $X_1, X_2=1, 0$, or if either $X_1, X_2=0, 1$, or $X_1, X_2=1, 1$ provided that relay X_1 was energized before X_2.

An inspection of the excitation and flow matrices shown in Table II indicates that a static hazard is present in the transition of states two⟷four.

Table III shows the reliability matrix for the circuit shown in Fig. 2 with respect to the flow matrix shown in Table II. The elements of the reliability matrix were calculated by using the rules of step 3 of the analysis procedure, together with the rules in the section "Static Hazards."

Synthesis

The probabilistic logical synthesis problem involves being given a description of a specified logic property, the statistical properties of the contacts of the relays available, and the lower bound for the set of defined elements of the reliability matrix.

The objective of the synthesis procedure is to synthesize a relay circuit subject to certain constraints. The design should realize the specified logic property so that if an analysis of the resulting circuit is made, no coefficient of the reliability matrix will be less than the a priori lower bound. For practical purpose, two other kinds of constraints will be placed upon the nature of the relays employed in the circuit to be designed. The first kind of constraint will have to do with the quality of the individual relays. The second will have to do with the total quantity of the equipment involved.

The purpose of the first kind of constraint is to facilitate the obtaining of some specific results without too much effort. This constraint will consist of applying postulate 2 and making B equal to $1-A$ in Fig. 1. The latter condition makes the defining channel for the contacts symmetrical, as shown in Fig. 4(A) and (B).

The purpose of the second kind of constraint is to narrow the class of possible solutions to the synthesis problem by means of an economy condition. The

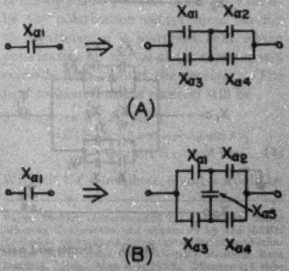


Fig. 5. A—Series-parallel network transformation. B—Bridge network transformation

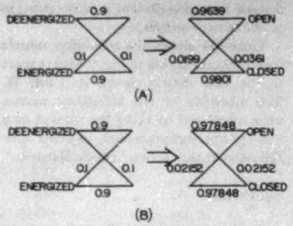


Fig. 6. Transformation of channels

A—Corresponding to Fig. 5(A)
B—Corresponding to Fig. 5(B)

economy condition will be aimed at either minimizing the number of contacts or the number of relays utilized in the circuit design.

The detailed procedure for undertaking the synthesis is as follows:

Step 1. From the specified logic properties of the problem, synthesize a minimal classical sequential circuit. If the classical circuit contains redundant secondary relays or contacts, in general, these will constitute an error producing redundancy. The classical circuit is to serve as a skeleton which is to be clothed with error reducing redundancy.

Step 2. Consider each normally open contact in the classical circuit from step 1. Assume that each of these contacts functions according to the statistics of the noisy symmetrical channel shown in Fig. 4(A). Consider each normally closed contact from step 1. Assume that each of these contacts functions according to the statistics of the noisy symmetrical channel shown in Fig. 4(B). Make an analysis of the resulting probabilistic circuit. Each element of the resulting reliability matrix will be a function f_{JK} of A_0.

Let α_0 ces be the specified lower bound for the reliability of the allowable state transitions. Then, $P_{JK}=f_{JK}(A_0)$. The problem is to find a greatest lower bound $A_{0M}=GLB(A_0)$ so that $f_{JK}(A_0)\geq\alpha_0$ ces for all J and K.

A lower bound to A_0 is sought so as to meet the reliability specifications with as crummy a set of relays as possible.[9] A greatest lower bound is sought so that the least amount of equipment need be used, consistent with the quality of the relays available for the construction of the circuit.

By postulate 2, if $A_{0M}<0.9$, then the circuit meets the reliability specifications without the need of using additional contacts to improve the reliability. If the defining probability A_{0A} for any contacts on the relays available for the synthesis are such that $0.9\leq A_{0M}<A_{0A}$, then no additional contacts are needed either.

Table IV. Flow Matrix (F) for Synthesis Example

	X_1X_2 00	01	11	10
Y	①	②	③	4
	1	2	⑤	④

F

Table V. Stochastic Matrix ($\mathfrak{M}$) for Circuit Shown in Fig. 7(B)

	1	2	3	4	5
1	..	A_0	..	A_0^2	..
2	A_0	..	1	..	..
3	..	1	..	A_0^4	..
4	A_0^3	..	..	..	A_0^6
5	..	A_0	..	A_0^3	..

$\mathfrak{M}$

Step 3. If $0.9\leq A_{0A}<A_{0M}$, then a local improvement in the reliability for these particular contacts can be made so as to meet the specified reliability for the overall circuit. For example, let $A_{0A}=0.90$, $A_{0M1}=0.95$, and $A_{0M2}=0.97$ for a particular contact denoted $x\alpha_1$. Then make either the series-parallel transformation shown in Fig. 5(A) or the bridge transformation shown in Fig. 5(B). Fig. 6(A) shows the transformation of channels corresponding to the series-parallel transformation. This is done for the case of a normally open contact only. Fig. 6(B) shows the transformation of channels corresponding to a bridge transformation. This is done for a normally open contact only.

Thus, as far as $A_{0M1}=0.95$, the series parallel and the bridge transformations will give the required local contact reliability. However, one contact may be saved by using the series-parallel transformation.

For the case in which $A_{0M2}=0.97$, a single application of the series-parallel transformation does not give sufficient reliability. However, the bridge transformation does give sufficient reliability.

The series-parallel transformation of a symmetric channel results in a nonsymmetrical channel. However, a bridge transformation of a symmetrical channel results in another symmetrical channel. This is because a bridge network is self-dual.

An alternative means of satisfying the specified reliability conditions for the circuit is to M-truplicate in the gross. By this is meant to synthesize a number of identical relay sequential circuits. None of these quite meet the required reliability specifications for the final network. However, by coupling a sufficient number

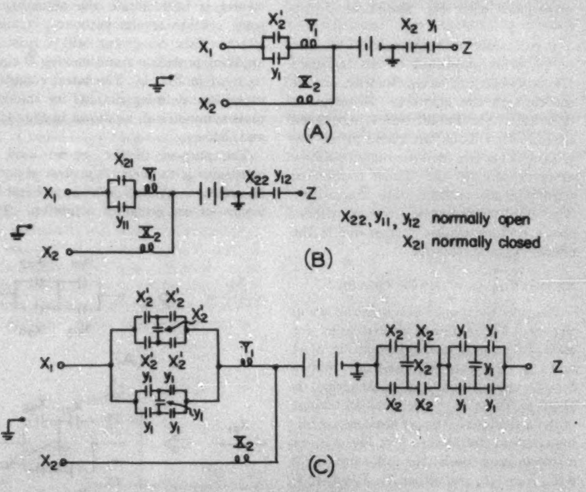


Fig. 7. Circuits and network for synthesis example

A—Corresponding classical circuit
B—Probabilistic circuit
C—Final network

of the identical circuits in the proper manner, an over-all circuit can be produced that will meet the required reliability specification.

Unlike the local improvement of reliability, the *M*-truplication of networks requires a large number of relays. If a considerable amount of local improvement is being considered, the spring load required for certain relays may exceed that which is present on the available relays. In such a case *M*-truplication in the gross gives a satisfactory solution to the synthesis problem. A majority relay, which is analogous to the majority organ described by Von Neumann, may be used as the coupling device for *M*-truplication in the gross.

Step 4. If a static hazard is present in the classical circuit of step 1 of the synthesis procedure, insert the missing prime implicant tie sets or cut sets. These inserted prime implicants will appear in the "skeleton" of the probabilistic circuit. To this extent they will be called probabilistic prime implicants or, more briefly, probimplicants. To decrease the probability of error due to the effects of static hazards, each of the probimplicants should be made as reliable as possible, consistent with the economic principle of "diminishing returns."

Let θ be the probability of error due to a malfunction in a particular probimplicant tie set or cut set. Let ϵ be the probability of error, for an allowable transition, due to the effects of a static hazard. The corrected probability of error ϵ_s, for an allowable state transition, due to the effects of a static hazard is $\epsilon_s=\theta\times\epsilon$.

SIMPLE EXAMPLE SHOWING SYNTHESIS PROCEDURES

Synthesize a relay sequential circuit with the logic property that the output line is to have the transmission function $Z=1$ if and only if the ordered set of inputs are such that $X_1, X_2=1, 1$, where X_1 became 1 before X_2. No reliability for an allowable transition of state in the resulting circuit is to be less than 3.12 ces. Assume that the available relays are defined by the channels shown in Fig. 4(A) and (B) with $A_0=0.9$.

Table IV is a flow table for the logic properties of the corresponding classical circuit. Fig. 7(A) shows a classical sequential circuit derived from inspecting the flow table according to the techniques developed by Huffman.[4] Fig. 7(B) presents the initial probabilistic sequential circuit derived from the corresponding classical circuit shown in Fig. 7(A).

The stochastic matrix for the analysis of the circuit shown in Fig. 7(B) is given in Table V. It should be noted that there is never any uncertainty in the state of the input line X_1.

Hence $-\log_2(1-A_0^4)\geq 3.12$, or $A_0=\geq 0.97$. Thus $A_{0M}=0.97$ since $A_{0A}=0.90$.

Local improvements in the individual contacts will be made. Refer to the bridge transformation shown in Fig. 5(B). This satisfies the needed local improvement in contact reliability as found by inspecting Fig. 6(B). It should be noted that the series-parallel transformation is inadequate. A final network is shown in Fig. 7(C).

References

1. PROBABILISTIC LOGICS AND THE SYNTHESIS OF RELIABLE ORGANISMS FROM UNRELIABLE COMPONENTS, John von Neumann. "Automata Studies," C. E. Shannon, J. McCarthy, editors. Princeton University Press, Princeton, N. J., 1956, pp. 43–98.

2. RELIABLE CIRCUITS USING LESS RELIABLE RELAYS, E. F. Moore, C. E. Shannon. *Journal*, Franklin Institute, Philadelphia, Pa., PART I, vol. 262, Sept. 1956, pp. 191–208; PART II, Oct., pp. 281–97.

3. A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY, W. S. McCulloch, W. Pitts. *Bulletin of Mathematical Biophysics*, Chicago, Ill., vol. 5, 1943, pp. 115–33.

4. THE SYNTHESIS OF SEQUENTIAL SWITCHING CIRCUITS, D. A. Huffman. *Journal*, Franklin Institute, PART I, vol. 257, Mar. 1954, pp. 161–90; PART II, Apr., pp. 275–303; see also reference 6.

5. THE DESIGN AND USE OF HAZARD-FREE SWITCHING NETWORKS, D. A. Huffman. *Journal*, Association for Computing Machinery, New York, N. Y., vol. 4, no. 1, Jan. 1957, pp. 47–62; see also reference 6.

6. SWITCHING CIRCUITS AND LOGICAL DESIGN (book), S. H. Caldwell. John Wiley & Sons, Inc., New York, N. Y., 1958.

7. THE MATHEMATICAL THEORY OF COMMUNICATION (book), C. E. Shannon, Warren Weaver. University of Illinois Press, Urbana, Ill., 1949.

8. STOCHASTIC PROCESSES, J. L. Doob. John Wiley & Sons, Inc., 1953, p. 97.

Measurement of the Scattering Matrix with an Intervening Ionosphere

HENRY BRYSK
NONMEMBER AIEE

WITH THE ADVENT of radar scattering from the moon and earth satellites, it is interesting to examine the effect of the interposition of a magnetoionic medium (the ionosphere) between the observer and the scatterer upon measurements of the scattering matrix. The integrated electron density of the ionosphere is of the order of 10^{13} electrons per centimeter squared, with diurnal variations of a factor of two. A typical value of the geomagnetic field is $^1/_2$ gauss. For radar frequencies, the Faraday effect may be considered as a rotation of the plane of polarization without absorption. At 200 megacycles, the rotation is about π each way. It varies as the inverse square of the frequency. For simplicity, the discussion is restricted to backscattering; no new essential features appear in the general case. The scattering matrix of interest is, of course, that which would be present in the absence of the ionosphere, not that due to the composite system of scatterer plus ionosphere. The complication is that it can be observed only with the effect of the Faraday rotation superimposed.

With the direction of propagation as the z-axis, and let $\vec{E}^i$ be the incident field and $\vec{E}^s$ the scattered field. The scattering matrix is then described by

$$\begin{aligned} E_x^s &= S_{11}E_x^i+S_{12}E_y^i \\ E_y^s &= S_{21}E_x^i+S_{22}E_y^i \end{aligned} \qquad (1)$$

where the S_{jk} are complex numbers. By reciprocity, $S_{12}=S_{21}$. An over-all absolute phase is irrelevant. There are thus five independent real numbers that determine the S matrix: The magnitudes of S_{11}, S_{22}, and S_{12}, and the relative phases between them.

Denote the electric polarization vector of the field incident on the scatterer by $(\hat{x}\cos\theta+\hat{y}\sin\theta)$, and let $(\hat{x}\cos\alpha+\hat{y}\sin\alpha)$ be the polarization vector upon leaving the scatterer of a scattered field whose polarization on arrival at the receiver will coincide with that of the receiver. The field measured at the receiver will be

$$\begin{aligned} E^r &= E_x^s\cos\alpha+E_y^s\sin\alpha \\ &= E^i(S_{11}\cos\theta\cos\alpha+S_{22}\sin\theta\times \\ &\qquad \sin\alpha+S_{12}\sin(\theta+\alpha)) \end{aligned} \qquad (2)$$

Without the ionosphere, θ and α would be

Paper 58-532, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958. Manuscript submitted January 24, 1958; made available for printing May 6, 1958.

HENRY BRYSK is with the University of Michigan, Ann Arbor, Mich.

the polarization angles of the transmitter and receiver respectively, and the S-matrix would be determined by making five measurements (e.g., for the pairs of values $(\theta=0,\ \alpha=0)$, $(\theta=\pi/2,\ \alpha=\pi/2)$, and $(\theta=0,\ \alpha=\pi/2)$, one obtains $(E^r/E^t)=S_{11}$, S_{22}, and S_{12} respectively, so that if one measures the amplitude of each of the three returns and the two relative phases, the S-matrix is determined).

With the ionosphere in the way, the Faraday effect rotates the polarization vector by an angle ψ, so that if the radiation is incident with a polarization angle θ it must have started with an angle $\phi=\theta-\psi$; similarly, to receive the component of the scattered field that leaves the scatterer with polarization angle α one must receive with a polarization angle $\beta=\alpha+\psi$. Thus, if the transmitter has a polarization $(\hat{x}\cos\phi+\hat{y}\sin\phi)$ and the receiver a polarization $(\hat{x}\cos\beta+\hat{y}\sin\beta)$, the measured field will be

$$\begin{aligned}E^r=E^t[&S_{11}\cos(\phi+\psi)\cos(\beta-\psi)+\\&S_{22}\sin(\phi+\psi)\sin(\beta-\psi)+\\&S_{12}\sin(\phi+\beta)]\\=E^t[&S_{12}\sin(\phi+\beta)+(1/2)\\&(S_{11}-S_{22})\cos(\phi+\beta)+\\&(1/2)(S_{11}+S_{22})\cos(\phi-\beta)\cos\\&2\psi+(1/2)(S_{11}+S_{22})\sin(\phi-\beta)\sin 2\psi]\end{aligned}\tag{3}$$

This expression contains functions of ϕ and β which are known for a particular choice of transmitter and receiver polarizations, and the unknown quantities S_{12}, $(S_{11}-S_{22})$, $(S_{11}+S_{22})\cos 2\psi$, and $(S_{11}+S_{22})\sin 2\psi$. Four pairs of polarization angle settings are now necessary, supplying seven measurements; these suffice to determine the amplitude and phase of the unknowns just mentioned. Since the trigonometric functions are real, there is a redundancy in that the phase of the last two unknowns is the same (modulo π). It is not proper simply to discard one phase measurement on this account; the corresponding amplitude is then a quadratic expression which possesses an extraneous root (unless one of the unknowns vanishes). One can, however, use delay line circuitry to superpose the returns without loss of phase before the recording stage in such a way that the resulting linear combinations are the above unknowns; after this is done, only six quantities need be recorded for maximum information about the scattering matrix (discarding the relative sign of $\cos 2\psi$ and $\sin 2\psi$ is of no importance).

These measurements do not, however, completely determine the scattering matrix. While $\cos 2\psi$ (or $\sin 2\psi$) is real, it can be either positive or negative; unless this is known, there is an ambiguity of π in the phase of $(S_{11}+S_{22})$. Because of this ambiguity, the measurements yield two scattering matrices to choose between, only one of which is, of course, correct. (The two are identical for a matrix proportional to the unit matrix, such as occurs for the sphere). If the two solutions are denoted by superscripts a and b, and the over-all phase of one with respect to the other is fixed, they are related by

$$S_{11}^a=S_{22}^b,\ S_{22}^a=S_{11}^b,\ S_{12}^a=-S_{12}^b\tag{4}$$

The two matrices can correspond to the same object, rotated through 90 degrees about the z-axis. The dichotomy may often not matter in practice.

The complete specification of the s-matrix requires either of two additional data:

1. A determination of the Faraday rotation adequate to ascertain the sign of $\cos 2\psi$. At high enough frequencies, 2ψ will be in the first quadrant, so there is no ambiguity. At lower frequencies, simultaneous observations at two frequencies that differ slightly can pinpoint the scale of the rotation.
2. Sufficient knowledge of the physical system under observation from other sources to be able to decide which of the two scattering matrices makes sense.

Magnetic Amplifier Bibliography 1951–1956

AIEE COMMITTEE REPORT

THIS bibliography covers domestic and foreign literature for the years 1951 through 1956. The patents listed cover those issued between January 1951 and July 1956. Earlier magnetic amplifier literature is covered in the bibliography prepared by J. G. Miles.[12]

The bibliography is divided into four major subdivisions as follows:

- I. English Language Articles
- II. Foreign Language Articles
- III. Domestic Patents
- IV. Foreign Patents

Within each major subdivision, papers and patents are classified according to subject matter as follows:

- A. *Theory, Analysis, and Design*
  Papers dealing primarily with the theory of operation of magnetic amplifiers, or with the analysis and design of circuits.
- B. *Magnetic Regulation and Servo Applications*
  Applications of magnetic amplifiers in the control of voltage, current, speed, temperature, frequency, position, etc.
- C. *Magnetic Frequency Conversion*
  Sub- and ultraharmonic generators and magnetic frequency converters.
- D. *Operational and Instrument Magnetic Amplifiers*
  Metering and operational magnetic amplifiers and applications in analog computation. Magnetic modulators.
- E. *Static Magnetic Switching Applications*
  Digital computer applications and discrete step output magnetic amplifiers. Bistate magnetic amplifiers.
- F. *Transistor–Magnetic Device Combinations*
  All applications where transistors are used in conjunction with magnetic devices to provide new capabilities.
- G. *Magnetism and Magnetic Materials*
  Magnetic phenomena and the properties of magnetic amplifier materials.
- H. *Measurements and Testing*
  Techniques for the measurement of magnetic properties, core matching, and grading. Methods for measuring magnetic amplifier performance.
- J. *Associated Devices*
  Components used in conjunction with saturable reactors in magnetic amplifier circuitry. Devices generally associated with magnetic amplifier technique such as rectifiers, reference elements, etc.
- K. *Ferroresonance and Its Applications*
  The phenomenon of ferroresonance and its applications in regulation and harmonic frequency generation.
- L. *Manufacturing Techniques, Environmental Effects, and Reliability*
  Manufacturing and packaging techniques utilized to obtain a product reliable over a wide range of environmental conditions. The effects of radiation and temperature on magnetic amplifier performance.

Each subject heading is further subdivided into years. For each year, the papers under a given subject classification are listed alphabetically by author, except for patents which are listed in order of increasing patent numbers.

A list of abbreviations of the journals referred to in the bibliography and a list of general references on magnetic amplifiers follows.

Paper 58-996, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the Joint AIEE-IRE Nonlinear Magnetics and Magnetic Amplifiers Conference, Los Angeles, Calif., August 6–8, 1958. Manuscript submitted April 3, 1958; made available for printing May 1, 1958.

The members of the Bibliography Working Group of the Applications Subcommittee of the AIEE Magnetic Amplifiers Committee are B. J. ALEXANDER, *chairman;* V. B. Boros, W. A. Geyger, R. E. Morgan, and R. A. Raber. Other members of the Applications Subcommittee made significant contributions to this bibliography. Special mention is due L. W. Buechler for his initial work which provided the basis for this bibliography, and W. G. Evans, David Katz, H. M. Ogle, H. F. Storm, and T. G. Wilson, all of whom contributed lists of references.

List of Journals and Their Abbreviations

ACEC Rev. (Belgium)	*ACEC Revue.* Ateliers de Constructions Électriques de Charleroi, Rue Vital Francoise, Marcinelle-Charleroi, Belgium
AEG Mitt. (Berlin)	*AEG Mitteilungen.* Hohenzollerndamm 150, Berlin-Grunewald, Germany
Aeron. Engg. Rev.	*Aeronautical Engineering Review.* Institute of the Aeronautical Sciences, Inc., 2 E. 64th Street, New York 21, N.Y.
AIEE Trans.	*Transactions,* American Institute of Electrical Engineers, 33 W. 39th Street, New York 18, N.Y.
Allis-Chalmers Elec. Rev.	*Allis-Chalmers Electrical Review.* Allis-Chalmers, Milwaukee 1, Wis.
Appl. Sci. Research (Hague)	*Applied Scientific Research.* Martinus Nijhoff, The Hague, Netherlands
Arch. Elektr. Übertr. (Stuttgart)	*Archiv der Elektrischen Übertragung.* S. Hirzel Verlag, Birkenwaldstrasse 185, Stuttgart, Germany
ASEA J.	*ASEA Journal.* ASEA Electric Ltd., Fulbourne Road, London, E. 17; Allmänna Svenska Elektriska AB, Vasteras, Sweden
ATM (Munich)	*Archiv für Technisches Messen.* Leibniz Verlag, Lotzbeckstr. 2b, Munich, Germany
Australian J. Appl. Science	*Australian Journal of Applied Science.* Commonwealth Scientific and Industrial Research Organization, 314 Albert Street, East Melbourne, C2, Australia
Automation and Remote Control (USSR)	*Avtomatika i Telemekhanika.* Institute of Automatics and Telemechanics, Moscow Center, Mal. Kharitonievsky, Per 4, Moscow, USSR
Blast F. & Steel Pl.	*Blast Furnace and Steel Plant.* Steel Publications, Inc., 330 Grant Street, Pittsburgh 30, Pa.
Brit. Commun. & Electronics	*British Communications and Electronics.* Drury House, Russell Street, Drury Lane, London, W.C. 2, England.
BSTJ	*The Bell System Technical Journal.* American Telephone and Telegraph Company, 195 Broadway, New York 7, N.Y.
Bull. de l'Association Suisse Electr.	*Bulletin de l'Association Suisse des Electriciens.* Secretariat General de l'A.S.E. et de l'U.C.S. Seefeldstrasse 301, Zurich 8, Switzerland
Bull. Int. Rly. Cong. Ass.	*Monthly Bulletin of the International Railway Congress Association* (English Edition). 19 rue de Beau-Site, Brussels, Belgium
Bull. Sci. Ass. Ing. Élect. Montefiore	*Bulletin Scientifique de l'Association des Ingénieurs Électriciens Sortis, de l'Institut Électrotechnique Montefiore.* 31 Rue Saint Gilles, Liege, Belgium

Bull. Soc. Belge Électr.	*Bulletin Société, Belge des Électriciens.* 8 rue Sabatier, Marcinelle, Belgium
Bull. Soc. Franc. Électr.	*Bulletin de la Société Francaise des Électriciens.* 8-14 Avenue Pierre Larousse, Malakoff, Seine, France
Cont. Engg.	*Control Engineering.* McGraw-Hill Book Company, Inc., 330 W. 42nd Street, New York 36, N.Y.
Conv. Rec. IRE.	*Convention Record,* Institute of Radio Engineers, 1 E. 79th Street, New York 21, N.Y.
Direct Current	*Direct Current.* Garraway Ltd., 11A Kensington Church Street, London, W. 8, England
Dtsch. Electrotechnik	*Deutsche Electrotechnik.* Verlag Technik, Unter den Linden 12, Berlin, N.W. 7, Germany
Elec. Engg.	*Electrical Engineering.* American Institute of Electrical Engineers, 33 West 39th Street, New York, N.Y.
Elec. Mfg.	*Electrical Manufacturing.* Gage Publishing Company, Inc., 1250 Sixth Avenue, New York 20, N. Y.
Elec. Rev.	*Electrical Review.* Dorset House, Stamford Street, London, S.E. 1, England
Elec. Times	*Electrical Times.* Sardinia House, London, W.C. 2, England
Électricien	*L'Électricien.* 92 Rue Bonaparte, Paris VI°, France
Électricité	*Électricité.* Editions Science et Industrie, 6 av. Pierre Premier de Serbie, Paris 16°, France
Electricity (USSR)	*Elektrichestvo.* Cherkasski Per, Moscow, USSR
Electronic Engg.	*Electronic Engineering.* 28 Essex Street, London, W.C. 2, England
Electronic Ind. & Tele-Tech.	*Electronic Industries and Tele-Tech.* Title changed from *Tele-Tech and Electronic Industries* Sept. 1956
Electronics	*Electronics.* McGraw-Hill Book Company, Inc., 330 W. 42nd Street, New York, N. Y.; Aldwych House, London, W.C. 2, England
Electrotechnika (Budapest)	*Electrotechnika.* V. Honved-u. 22, Budapest, Hungary
Elektrotechnik (Prague)	*Elektrotechnik.* Vocelova 3, Prague XII, Czechoslovakia
Elektrotech. Obz.	*Elektrotechniky Obzor.* Vocelova 3, Prague XII, Czechoslovakia
Elektroteknikeren (Copenhagen)	*Elektroteknikeren.* Landskronagade 70, Copenhagen Ø, Denmark
Elektrotek T. (Oslo)	*Elektroteknisk Tidsskrift.* Fuglehauggt 11, Oslo, Norway
Elettrotecnica (Milan)	*L'Elettrotecnica.* Associazione Elettrotecnica Italiana, Via S. Paolo 10, Milan, Italy
Elliott J.	*Elliott Journal.* Elliott Brothers, Century Works, Lewisham, London, S.E. 13, England
El World	*Electrical World.* McGraw-Hill Book Company, Inc., 330 W. 42nd Street, New York, N. Y.; Aldwych House, London, W.C. 2, England
EM (Vienna)	*Elektrotechnik und Maschinenbau.* Springer-Verlag. Molkerbastei 5, Vienna I, Austria
Engg. Bull. (Siemens)	*Engineering Bulletin.* Siemens Brothers and Company, Ltd., Woolwich, London, S.E. 18, England
Engg. Journal	*The Engineering Journal.* Engineering Institute of Canada, 2050 Mansfield Street, Montreal, Que., Canada
Engineering	*Engineering.* 35-36 Bedford Street, London, W.C. 2, England
Engrs. Dig.	*Engineers' Digest.* Engineers' Digest, Ltd., 120 Wigmore St., London W. 1, England
ETZ (Berlin)	*Elektrotechnische Zeitschrift.* Verband Deutscher Elektrotechniker, Friedrich-Ebert-Str. 111, Wuppertal-Elberfeld, Germany
Frequenz (Berlin)	*Frequenz.* Schiele & Schon, Leuschnerdamm 13, Berlin, S.O. 36, Germany
Funk und Ton	*Funk und Ton.* Verlag für Radio-Foto-Kinotechnik, Eichborndamm 141/167, Berlin-Borsigwalde, Germany
GEC J.	*G.E.C. Journal.* General Electric Company, Magnet House, Kingsway, London, W.C. 2, England
GEC Telecomm.	*G.E.C. Telecommunications,* Coventry, England
GE Rev.	*General Electric Review.* General Electric Company, Schenectady 5, N. Y.
Indian J. Phys.	*Indian Journal of Physics* (and *Proceedings* of the Indian Association for the Cultivation of Science). 2 & 3 Lady Willingdon Road, Calcutta, India
Ingenieur (Hague)	*De Ingenieur.* Koninklijk Instituut van Ingenieurs, 23 Prinsessegracht, The Hague, Netherlands
Instruments	*Instruments.* Instruments Publishing Company, Inc., 1600 N. Main St., Pontiac, Ill.
Instrum. & Automation	Title extended from *Instruments,* 1954
Iron and Steel Engr.	*Iron and Steel Engineer.* Association of Iron and Steel Engineers, 1010 Empire Building, Pittsburgh 22, Pa.
J. Amer. Soc. Nav. Engrs.	*Journal,* American Society of Naval Engineers, Continental Building, 1012–14th Street, N. W., Washington 5, D.C.
J. Applied Physics	*Journal of Applied Physics.* American Institute of Physics, 335 E. 45th Street, New York 17, N. Y.
J. Franklin Inst.	*Journal,* Franklin Institute, 20th and the Parkway, Philadelphia 3, Pa.
J. Sci. Inst.	*Journal of Scientific Instruments.* Institute of Physics, 47 Belgrave Square, London, S.W. 1, England
J. Tech. Phys. (USSR)	*Journal of Technical Physics.* Izdatel'stvo Akademii Nauk SSSR, Moscow, USSR.
Machine Design	*Machine Design.* Penton Publishing Company, Penton Building, Cleveland 13, Ohio
Mesures (Paris)	*Mesures et Controle Industriele,* Comité d'Editions Techniques, 79 av. des Champs-Elysées, Paris 8°, France
Metropolitan-Vickers Gaz.	*Metropolitan-Vickers Gazette.* Metropolitan Vickers Electrical Company, Trafford Park, Manchester 17, England
Mining Elect. Mech. Engr.	*Mining Electrical and Mechanical Engineer.* Cromwell Buildings, Blackfriars Street, Manchester 3, England
Nachrichtentechnik (Berlin)	*Nachrichtentechnik.* VerlagTechnik, Unter den Linden 12, Berlin, N.W. 7, Germany
OTS	*Office of Technical Services, Reports.* Department of Commerce, Washington 25, D. C.
Philips Res. Rep.	*Philips Research Reports.* Century House, Shaftesbury Avenue, London, W.C. 2, England; Philips' Gloeilampenfabrieken, Eindhoven, Holland
Post Office El. Engrs. J.	*The Post Office Electrical Engineers' Journal.* Engineer-in-Chief's Office, Alder House, Aldersgate Street, London, E.C. 1, England
Power Engineering	*Power Engineering.* Technical Publishing Company, Barrington, Ill.
Proc. IEE	*Proceedings,* Institution of Electrical Engineers, Savoy Place, London, W.C. 2, England
Proc. IRE	*Proceedings,* Institute of Radio Engineers, 1 E. 79th Street, New York 21, N. Y.
Proc. NEC	*Proceedings,* National Electronics Conference, Inc., 84 E. Randolph Street, Chicago, Ill.

Proc. STC on Mag. Amp., T-86	*Proceedings*, Special Technical Conference on Magnetic Amplifiers, *AIEE Special Publication T-86*, 1956. American Institute of Electrical Engineers, 33 W. 39th Street, New York 18, N. Y.
Prod. Engg	*Product Engineering*. McGraw-Hill Book Publishing Company, 330 W. 42nd Street, New York 18, N. Y.
Rad. & TV News	*Radio & TV News*. Ziff-Davis Publishing Company, 64 E. Lake Street, Chicago 1, Ill.
Radio-Electronic Engg	*Radio-Electronic Engineering* (published as a special edition in a limited number of copies of *Radio & TV News*)
Radio-Electronics	*Radio-Electronics*. 154 W. 14th Street, New York 11, N. Y.
Radio Engineering (USSR)	*Radiotekhnika*, ul. Kirova, 40, Moscow, USSR
Radio Francaise	*La Radio Francaise*. 92 Rue Bonaparte, Paris VI^e, France
RCA Rev.	*RCA Review*. Radio Corporation of America, Princeton, N. J.
Regelungstechnik (Munich)	*Regelungstechnik*. (article reprinted in *Entwicklungsber*, Siemens and Halske, Munich, Germany)
Research & Engg	*Research and Engineering*. 103 Park Avenue, New York 17, N. Y.
Rev. Gen. de l'Électricité	*Revue Générale d'Électricité*. 12, Place Henri-Bergson, Paris VIII^e, France
R. Sci. Inst.	*Review of Scientific Instruments*. American Institute of Physics, 335 E. 45th Street, New York 17, N. Y.
Siemens-Zs	*Siemens-Zeitschrift*. Siemens-Halske, Werner-Von Siemens-Str. 50, Erlangen, Germany
Slaboproudy Obzor	*Slaboproudy Obzor*. Krabovska ulice 8, Prague II, Czechoslovakia
Soc. Instrument Tech. Trans.	*Society of Instrument Technology, Transactions*. Hill House, Seal Hollow Road, Sevenoaks, Kent, England
Tele-Tech	*Tele-Tech*. Chestnut and 56th Streets, Philadelphia, Pa.
Tele-Tech & Elec. Ind.	Title extended from *Tele-Tech*, Jan. 1954
Tijdschr. (Amsterdam)	*Tijdschrift*, Van het Nederlandsch Radiogenootschap, Oude Outrechtscheweg 8, Baarn, Amsterdam, Holland
Trans. PG, IRE	*Transactions*, Professional Groups, Institute of Radio Engineers, 1 E. 79th Street, New York 21, N. Y.
TSF et TV	*TSF et TV*. 40 Rue de Seine, Paris 6^e, France
USNOL Reports	*United States Naval Ordnance Laboratory Reports*. Washington, D. C.
USNRL Reports	*United States Naval Research Laboratory Reports*. Washington, D. C.
VDE-Fachberichte	*Verband Deutscher Elektrotechniker Fachberichte*. VDE-Verlag, Friedrich-Ebert-Strasse 111, Wuppertal-Elberfeld, Germany
VDI Zs	*VDI Zeitschrift*. Verein Deutscher Ingenieure-Verlag, Bongard Strasse 3, Düsseldorf, Germany
Westinghouse Engr	*The Westinghouse Engineer*. P. O. Box 2278, 3 Gateway Center, Pittsburgh 30, Pa.
Wireless Engr	*Wireless Engineer*. Dorset House, Stamford Street, London, S.E. 1, England
Wireless World	*Wireless World*. Dorset House, Stamford Street, London, S.E. 1, England
Z. angew. Math. Phys. (Basle)	*Zeitschrift für angewandte Mathematik und Physik*. Verlag Birkhauser A. G., Basle 10, Switzerland

General References on Magnetic Amplifiers

1. Saturating Core Devices (book), L. R. Crow. Scientific Book Publishing Company, Vincennes, Ind., 1949.

2. Magnetic Amplifier (book), G. M. Ettinger. Methuen, London; John Wiley & Sons, Inc., New York, N. Y., 1953, 88 pp.

3. Magnetic Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., 330 W. 42nd. St., New York 36, N. Y., second edition, 1957, 277 pp.

4. The Transductor Amplifier (book), U. H. Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947, 178 pp.

5. The Transductor (book), A. U. Lamm. Esselte Aktiebolaget, Stockholm, Sweden, second edition, 1948.

6. Transductors and Magnetic Amplifiers (book), A. G. Milnes. Macmillan & Company, Ltd., London, 1957, 286 pp.

7. The Magnetic Amplifier (book), J. H. Reyner. Stuart and Richards, London, 1950, 119 pp.

8. Magnetic Amplifiers (in Russian), M. A. Rozenblat. Moscow, USSR, 1956, 824 pp.

9. Magnetic Amplifiers and Saturable Reactors (book), M. G. Say, editor. George Newnes, London, 1954, 200 pp.

10. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, 545 pp.

11. Bibliography on Transductors and Magnetic Amplifier, H. B. Rex. *Instruments*, vol. 21, no. 4, 1948.

12. Bibliography of Magnetic Amplifiers Devices and Saturable Reactor Art, J. G. Miles. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 2104–23 (901 references).

13. Bibliography on Magnetic Amplifiers for 1951–1954. *Automation and Remote Control* (USSR), vol. 17, no. 5, May 1956, pp. 471–87.

14. Bibliography on Magnetic Amplifiers for 1955, G. V. Subbotina. *Ibid.*, no. 9, Sept. 1956, pp. 858–64.

I: English Language Articles

A. Theory, Analysis, and Design

1951

15. Overvoltages in Saturable Series Devices, A. Boyajian, G. Camilli. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 1845–53.

16. The Magnetic Amplifier, A. A. Basshart. *Radio Electronics*, vol. 22, no. 12, 1951, pp. 40–42.

17. Low Input-Power Level Magnetic Amplifiers, L. W. Buechler. *Proc. NEC*, vol. 7, 1951, pp. 254–59.

18. Saturable Reactors and Control Oscillators, E. Bukstein. *Rad. & TV. News*, vol. 45, Apr. 1951, pp. 42, 120–121.

19. Analysis and Design of Self-Saturable Amplifiers, S. B. Cohen. *Proc. IRE*, 1951, pp. 1009–20.

20. Applying Magnetic Amplifiers, W. J. Dornhoefer, V. Krummenacher. *Elec. Mfg.*, vol. 47, no. 3, 1951, p. 94; no. 4, p. 112; vol. 48, no. 2, p. 106; no. 3, p. 92.

21. The Steady State and Transient Analysis of the Doubler and Full-Wave Magnetic Amplifier with Free Excitation for "Rectangular" Hysteresis Core Material and Resistive Load, A. Erkman. *Report R-260-51, PIB-201, M.R.I.*, Polytechnic Insitute of Brooklyn, Brooklyn, N. Y., 1951.

22. Steady State Analysis of Self-Saturating Magnetic Amplifiers Based on Linear Approximations of the Magnetization Curves, W. H. Esselman. *AIEE Trans.*, vol. 70, pt. I, 1951, pp. 451–54.

23. A Magnetic Amplifier of High Input Impedance, G. M. Ettinger. *Proc. NEC*, vol. 7, 1951, pp. 523–28.

24. Parallel-Connected Magnetic Amplifiers, J. D. Fahnestock. *Electronics*, vol. 24, no. II, 1951, pp. 140, 224.

25. High Gain Magnetic Amplifier. Theory of the Self-Excited Transductor, R. Feinberg. *Wireless Engr.*, vol. 28, no. 332, May 1951, pp. 151–55.

26. Rules of Similitude for Magnetic Amplifier Systems, L. A. Finzi. *Proc. NEC*, vol. 7, 1951, pp. 498–514.

27. The Transient Response of Magnetic Amplifiers Cases of Negligible Commutation, L. A. Finzi, D. P. Chandler, D. C. Beaumariage. *AIEE Trans.*, vol. 70, pt. I, 1951, pp. 934–42.

28. Magnetic Amplifiers of the Balance Detector Type—Their Basic Principles, Characteristics and Applications, W. A. Geyger. *Ibid.*, pt. II, pp. 1707–20.

29. Design of Saturable Reactors, J. E. Hart. *OTS, PB 110337*, 1951.

30. Use of the Flux-Charge Concept in Magnetic Amplifier Analysis, P. M. Kintor. *Proc. NEC*, vol. 7, 1951, pp. 529–36.

31. Predetermination of Control Characteristics of Half-Wave Self-Saturated Magnetic Amplifiers, H. Lehmann. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 2097–2103.

32. Some General Properties of Magnetic Amplifiers, J. M. Manley. *Proc. IRE*, vol. 39, no. 3, Mar. 1951, pp. 242–51.

33. The Behavior of a Series-Transductor Magnetic Amplifier with Directly-Connected or Rectifier-Fed Loads, A. G. Milnes. *Proc. IEE, Paper no. 1190*, vol. 98, Measurements Section, pp. 18–23, ref. 10.

34. Steady-State and Transient Analysis of an Idealized Series-Connected Magnetic

I: English Language Articles

A. Theory, Analysis, and Design

1951 (cont'd.)

Amplifier, L. A. Pipes. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 2129–33.

35. On the Mechanics of Magnetic Amplifier Operation, R. A. Ramey. *USNRL Report no. 3799*, Jan. 22, 1951.

36. On the Control of Magnetic Amplifiers, R. A. Ramey. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 1214–23.

37. The Application of a 60-Cycle Magnetic Amplifier to a Position Indicating Servomechanism, A. E. Schmid. *Proc. NEC*, vol. 7, 1951, pp. 515–22.

38. How Magnetic Amplifiers Operate, C. H. Stearley. *Prod. Engg.*, 1951, vol. 22, no. 1, pp. 132–36.

39. Transient Response of Saturable Reactors with Resistive Load, H. F. Storm. *AIEE Trans.*, vol. 70, pt. I, 1951, pp. 95–102.

40. Magnetic Amplifiers: A Rising Star in Naval Electronics, A. Vincent. *J. Amer. Soc. Nav. Engrs.*, vol. 63, no. 4, Nov. 1951, pp. 903–907.

41. Study of Fundamentals and Characteristics of Magnetic Amplifiers for Flight Control Systems, J. R. Walker, D. G. Wilson. *U.S. Air Force Tech. Report no. 5881*, Part I, Design and Analysis of Basic Circuits, Dec. 1951.

42. Basic Operating Principles of Magnetic Amplifiers, E. V. Weir. *Proc. NEC*, vol. 7, 1951, pp. 235–40.

1952

43. The Design of a Practical D-C Amplifier Based on the Second-Harmonic Type of Magnetic Modulation, P. J. Baxandall, S. W. Noble. *Proc. IEE.*, vol. 99, no. 70, pt. II, 1952, pp. 327–48.

44. Some Aspects of Magnetic Amplifier Technique, F. E. Butcher, R. Willheim. *Proc. IRE*, vol. 40, 1952, pp. 261–70.

45. The Figure of Merit of a Magnetic Amplifier, J. T. Carleton, W. F. Horton. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 239–45.

46. Magnetic Amplifier Definitions, W. H. Elliot. *Elec. Mfg.*, July 1952, p. 88.

47. The Effective Feedback Ratio of Magnetic Amplifiers, L. A. Finzi. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 157–64.

48. Comparison of Methods of Analysis of Magnetic Amplifiers, L. A. Finzi, G. F. Pittman, Jr. *Proc. NEC*, vol. 8, 1952, pp. 144–57.

49. Compensating for the Quiescent Current in Multi-Stage Magnetic Amplifiers, A. S. Fitzgerald. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 7–12.

50. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, E. Calvin Johnson. *Ibid.*, pt. II, July, pp. 169–81.

51. A High Speed Magnetic Amplifier, L. J. Johnson. *Elec. Mfg.*, vol. 50, no. 5, 1952, p. 98.

52. A Study of Amplifier Saturation and Magnetic Saturation in a Servomechanism, C. Leondes. *AIEE CP 52-198* (not available).

53. A Simple Magnetic Amplifier, E. Levy. *Radio-Electronic Engg.*, vol. 23, no. 9, June 1952, p. 56.

54. Types of Magnetic Amplifiers, J. G. Miles. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 229–38.

55. The Behavior of a Series Transductor Magnetic Amplifier with Directly Connected or Rectified Loads, A. G. Milnes. *Proc. IEE*, vol. 99, pt. II, 1952, pp. 13–23.

56. Three-Phase Transductor Circuits for Magnetic Amplifiers, A. G. Milnes. *Ibid.*, pt. IV, 1952, pp. 336–57.

57. A Mathematical Analysis of an Inductively Loaded Parallel-Connected Magnetic Amplifier, L. A. Pipes. *J. Applied Physics*, vol. 23, no. 12, Dec. 1952.

58. The Design of Single Phase Transductors, H. A. Ross. *Australian J. Appl. Science*, vol. 3, Dec. 1952, pp. 277–92.

59. The Effect of Commutation on the Stability of Magnetic Amplifiers, M. Sakamoto. *AIEE CP 52-223* (available on request).

60. Saturable Reactors with Inductive D-C Load, Pt. I. Steady-State Operation, H. F. Storm. *AIEE Trans.*, vol. 71, pt. I, Nov. 1952, pp. 335–43.

61. Basic Operating Principles of Magnetic Amplifiers, E. V. Weir. *J. Radio Comm.*, Great Barrington, Mass., 1952, no. 1, pp. 13–14.

62. Series-Connected Magnetic Amplifier with Inductive Loading, T. G. Wilson. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 101–110.

1953

63. Note on the Optimum Input Winding Resistance of a Magnetic Amplifier Employing Voltage Feedback, P. D. Atkinson. *Elliott J.*, vol. 1, May 1953, pp. 102–03.

64. Magnetic Amplifier with Reset Control, G. M. Attura. *Electronics*, vol. 26, June 1953, pp. 161–63.

65. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, S. B. Batdorf, W. N. Johnson. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 223–27.

66. The Figure of Merit of Magnetic Amplifiers, J. T. Carleton, W. F. Horton. *Elec. Engg.*, vol. 72, Mar. 1953, p. 229.

67. An Approximate Graphical Analysis of the Steady State Response of Non-Linear Networks, S. Duinker. *Philips Res. Rep.*, vol. 8, Apr. 1953, pp. 133–47.

68. Methods of Analysis of Magnetic Amplifiers, L. A. Finzi, G. F. Pittman, Jr. *Elec. Engg.*, vol. 72, no. 8, 1953, pp. 690–94.

69. The Study of Magnetic Inverters for Amplification of Low Input Power D-C Signals, E. H. Frost-Smith. *Proc. IEE*, vol. 100, pt. II, no. 76, 1953, pp. 362–375.

70. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 383–95.

71. A Half-Wave Auto-Transformer Magnetic Amplifier, E. T. Hooper. *USNOL Report 2719*, 1953.

72. Flux Preset High-Speed Magnetic Amplifiers, C. B. House. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 728–35.

73. The Magnetic Amplifier, W. C. Johnson. *Electronic Engg.*, vol. 72, no. 7, 1953, pp. 583–88.

74. The Influence of Magnetic Amplifier Circuits Upon the Operating-Hysteresis Loops, H. W. Lord. *AIEE Trans*, vol. 72, pt. I, 1953, pp. 721–28.

75. The Amplistat or Magnetic Amplifier, W. Miller. *Engrs. Dig.*, vol. 14, no. 17, July 1953, pp. 260–62.

76. Magnetic-Amplifier Circuits and Applications, R. A. Ramey. *Elec. Engg.*, vol. 72, no. 9, 1953, pp. 791–95.

77. Magnetic Amplifiers, R. W. Roberts. *Machine Design*, vol. 25, no. 24, Apr. 1953, pp. 171–79.

78. Magnetic Converter D-C Amplifier, W. A. Rote. *Electronics*, vol. 26, no. 12, 1953, pp. 170–73.

79. Commutation Effect on Magnetic Amplifier Stability, M. Sakamoto. *Elec. Engg.*, no. 2, 1953, p. 165.

80. Fast Response with Magnetic Amplifiers, D. G. Scorgie. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 741–49.

81. Saturable Reactors with Inductive D-C Load. Pt. II. Transient Response, H. F. Storm. *Ibid.*, pp. 182–92.

82. Theory of Magnetic Amplifiers with Square-Loop Core Materials, H. F. Storm. *Ibid.*, pp. 629–40.

83. Preliminary Analysis of the Half-Wave Bridge Magnetic Amplifier, H. H. Woodson. *USNOL Report 3506*, 1953.

1954

84. An Even-Harmonic Magnetic Amplifier and Some Applications to Measurement and Control, P. D. Atkinson, A. V. Hemingway. *Electronic Engg.*, vol. 26, no. 11, Nov. 1954, pp. 482–86.

85. Characteristics of Magnetic Amplifiers for Industrial Use, R. G. Beadle, B. P. Chausse. *Elec. Engg.*, vol. 73, no. 11, 1954, pp. 1023–27.

86. General Properties of Electromagnetic Amplifiers, D. A. Bell. *Wireless Engr.*, vol. 31, no. 12, Dec. 1954, pp. 310–19.

87. Parallel-Connected Magnetic Amplifier, S. H. Chow. *J. Applied Physics*, vol. 25, no. 2, Feb. 1954, pp. 216–21.

88. On a Non-Linear Diffusion Equation Applied to the Magnetization of Saturable Reactors, S. H. Chow. *Ibid.*, no. 3, Mar., pp. 377–81.

89. Magnetic Amplifiers, J. F. Coales. *Proc. IEE*, vol. 101, pt. II, no. 80, 1954, pp. 83–99.

90. Alteration of the Dynamic Response of Magnetic Amplifiers by Feedback, R. O. Decker. *AIEE Trans.*, vol. 79, pt. 1, 1954, pp. 658–65.

91. Negative Inductance Cuts Magnetic-Amplifier Lag, G. M. Ettinger. *Electronics*, vol. 27, no. 1, 1954, pp. 162–63.

92. Magnetic Amplifiers, G. M. Ettinger. *Elec. Rev.*, vol. 155, no. 9, 1954, pp. 348–52.

93. The Analysis of Magnetic Amplifiers with Various Types of Load. I: Load Currents for Given Angle of Firing, L. A. Finzi, R. R. Jackson. *AIEE Trans.* vol. 73, pt. I, July 1954, pp. 270–78.

94. The Operation of Magnetic Amplifiers with Various Types of Load. II: Controlling the Angle of Firing. The Transfer Characteristics of Amplifiers with Low-Control Impedance, L. A. Finzi, R. R. Jackson. *Ibid.*, pp. 279–88.

95. Analysis of a Single-Core Magnetic Amplifier with Real Rectifiers and Core Functions, M. Frank, S. Rabotnick, J. R. Walker. *Ibid.*, pp. 182–88.

96. Basic Facts About Magnetic Amplifiers. H. A. Goldsmith, H. Herz, B. J. O'Neil. *Cont. Engg.*, vol. 1, no. 4, 1954, pp. 40–47.

97. A Two-Stage Low-Level Magnetic Amplifier with Negative Voltage Feedback, F. Gourash. *Proc. NEC*, vol. 10, 1954, pp. 146–55.

98. Self-Excited Series Transductors. Performance Under Various Circuit Conditions, D. W. Harvey. *Engineering*, vol. 178, no. 4626, 1954, pp. 399–403.

99. Fast Response Magnetic Servo Amplifier, F. Hill, J. A. Fingerett. *Electronics*, vol. 27, no. 10, 1954, pp. 170–73.

100. Flux Preset High-Speed Magnetic Amplifiers, C. B. House. *AIEE Trans.*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 728–35.

101. Fast Response Magnetic Amplifiers, G. E. Hughes, H. A. Miller. *Ibid.*, vol. 73, pt. I, Mar. 1954, pp. 69–75.

102. The Transductor and Its Applications, J. A. Hunt, R. H. Kelsall. *Metropolitan-Vickers Gas.*, vol. 25, no. 417, Apr. 1954, pp. 228–35.

103. Analysis of Magnetic Amplifiers by Difference Equations, P. R. Johannessen. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 700–11.

104. How to Determine Response Time of Magnetic Amplifiers, L. J. Johnson, S. E. Rauch. *Elec. Mfg.*, vol. 54, no. 4, 1954, pp. 90–94.

105. Single Phase Magnetic Amplifiers with Decicyle Response, L. J. Johnson, S. E. Rauch. *Proc. NEC*, vol. 10, 1954, pp. 162–67.

106. Instability in the Full-wave Self-Saturating Magnetic Amplifier with Inductive Load, H. L. Leon, A. B. Rosenstein. *Report 54-48*, Univ. of California, Los Angeles, Calif., 1954.

107. Magnetic Amplifier Circuits with Full-Wave Output and Half-Wave Control Signals, H. W. Lord. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 265–70.

108. Design Considerations of Half-Wave Bridge Magnetic Amplifiers, C. W. Lufcy, H. H. Woodson. *Ibid.*, pp. 220–26.

109. Transductor and Its Applications, J. A. Lunt, R. H. Kelsall. *Metropolitan-Vickers Gaz.*, vol. 25, no. 417, 1954, pp. 228–35.

110. Three Phase High-Speed Magnetic Amplifiers, A. E. Maine. *Electronic Engg.*, vol. 26, no. 315, 1954, pp. 180–85.

111. Auto-Self-Excited Transductors and Push-pull Circuit Theory, A. G. Milnes, T. S. Law. *Proc. IEE*, vol. 101, pt. II, no. 84, pp. 643–62.

112. Transductors with Four Limbed Cores, A. G. Milnes. *Ibid.*, no. 83, pp. 554–58.

113. The Biased Rectifier Amplifier-A Pulse Magnetic Amplifier, R. E. Morgan. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 584–90.

114. Pulse Relaxation Amplifier-A Low Level DC Magnetic Amplifier, R. E. Morgan, J. B. McFerran. *Ibid.*, pp. 245–49.

115. Single Ended Saturable Reactor Circuit with Quiescent Current Compensation, R. J. Radus. *Ibid.*, pp. 597–603.

116. Derivative Controlled Magnetic Amplifiers, A. D. Schnitzler. *Ibid.*, pp. 569–76.

117. Magnetic Amplifiers with Inductive D. C. Load, H. F. Storm. *Ibid.*, pp. 466–75.

1955

118. New Full-Wave Magnetic Amplifier Output Stage, P. W. Barnhart. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 139–41.

119. An Analysis of Optimum Core Configuration for Magnetic Amplifiers Using a Simplified Approach, B. D. Bedford, C. H. Willis, G. C. Dodson. *Ibid.*, pp. 62–70.

120. Figure of Merit in Magnetic Amplifiers, J. T. Carleton, W. F. Horton. *Tele-Tech & Elec. Ind.*, vol. 14, June 1955, pp. 96–97.

121. High-Frequency Operation of Self-Saturating Amplifiers, W. H. Collins. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 500–05.

122. Alteration of the Dynamic Response of Magnetic Amplifiers by Feedback, R. O. Decker. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 658–65.

123. Magnetic Amplifiers in Bistable Operation, L. A. Finzi, G. C. Feth. *Ibid.*, pt. 1, 1955, pp. 592–98.

124. Elimination of Asymmetry Zero-Drift Errors In Magnetic Servo Amplifiers, W. A. Geyger. *Ibid.*, pp. 239–44.

125. The Application of Perpendicularly Superposed Magnetic Fields, R. A. Heartz, H. Buelteman, Jr. *Ibid.*, pp. 655–59.

126. Theory of Magnetic Cross Valves, W. H. Higa. *Ibid.*, pp. 204–08.

127. Full-wave Half-Cycle Response Magnetic Amplifiers, C. B. House. *Ibid.*, pp. 541–52.

128. Newer Types of Magnetic Amplifiers, R. R. Jackson. *Engg. Journal*, vol. 38, no. 5, May 1955, pp. 635–39.

129. Fast Response Magnetic Servo Amplifier, J. W. Kallander. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 49–54.

130. The Residual Time Constant of Self-Saturating (Auto-Excited) Transductors, U. Krabbe. *Proc. IEE, Monograph no. 140 M*, July 1955.

131. Inductive Load Instability in Amplifiers, H. I. Leon, A. B. Rosenstein. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 8–20.

132. Magnetic Amplifier Design-Practical Approach, M. Lilienstein. *Elec. Mfg.*, Mar. 1955, pp. 90–98.

133. Survey of Magnetic Amplifiers, C. W. Lufcy. *Proc. IRE*, vol. 43, Apr. 1955, pp. 404–13.

134. A Comparison of the Bridge and Differential Half-Wave Magnetic Amplifier Circuits, McMorrow. *USNOL Report 3855*, 1955.

135. Influence of ID/OD Ratio on Static and Dynamic Magnetic Properties of Toroidal Cores, R. W. Roberts, R. I. Van Nice. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 599–607.

136. 160,000 Ampere High-Speed Magnetic Amplifier Design, A. B. Rosenstein. *Ibid.*, pp. 90–97.

137. Magnetic Amplifiers and Their Applications, R. J. Russell-Bates. *Soc. Instrument Tech. Trans.*, vol. 7, no. 4, Dec. 1955, pp. 152–62.

138. An All Magnetic Audio Amplifier System, J. J. Suozzi, E. T. Hooper. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 297–300.

139. Circuit Equations for Rectifier and Magnetic Amplifier Circuits, D. H. Schaefer. *USNRL Report 4567*, 1955.

140. Formulation of Magnetic Amplifier Circuit Equations, D. H. Schaefer. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 619–27.

1956

141. Half-Wave Magnetic Amplifier Behavior, J. Baranowski. *Technical note no. EL. 119*, Royal Aircraft Establishment, England, May 1956, 38 pp.

142. Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier, T. Bernstein, N. L. Schmitz. *AIEE Trans.*, vol. 75, pt. I, Sept. 1956, pp. 433–38.

143. Magnetic Amplifiers Serve to Teach Power Modulators, H. C. Bourne, Jr., R. M. Saunders. *Elec. Engg.* vol. 75, Oct. 1956, p. 943.

144. Dynamic Operation of Magnetic Amplifiers for Feedback Control Systems, H. C. Bourne, Jr., T. T. Kadota, D. Nitzan. *Scientific Rept. no.* 3, Institute of Engineering Research, U. of California, Berkeley, Calif., Sept.–Nov. 1956, 24 pp.

145. Core and Lamination Shape Factors in Magnetic Amplifier Design, H. R. Brownell. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 45.

146. Analysis of Magnetic Control Circuit, C. H. R. Campling. *Engg. Journal*, vol. 39, no. 10, Oct. 1956, pp. 1360–65.

147. Capacatively Coupled Magnetic Amplifiers, H. W. Collins. *Proc. STC. on Mag. Amp., T-86*, Apr. 1956, p. 211.

148. An Analysis of Magnetic Amplifiers in Push-Pull, D. L. Critchlow. *Ibid.*, p. 52.

149. Behavior of Saturable Reactors in Magnetic Amplifiers, P. N. Das. *Indian J. Phys.*, vol. 30, Mar. 1956, pp. 129–42.

150. Analysis of the Doubler Magnetic-Amplifier Circuit with Simulated Eddy-Current Effect, D. Doody. Microwave Research Inst., Polytechnic Inst. of Brooklyn, Sept. 1956, 28 pp.

151. Dynamic Behavior of 3-Phase Magnetic Amplifier, F. L. Ellert. *AIEE Trans.*, vol. 75, pt. I, Nov. 1956, pp. 554–60.

152. The Operation of the Self-Balancing Magnetic Amplifier, E. T. Hooper, A. D. Krall. *USNOL Rept. 4190*, 1956; also *AIEE Trans.*, *Ibid.*, pp. 79–84.

153. A General Introduction to the Analysis of Magnetic Amplifiers, W. C. Johnson. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 9.

154. Existence of an Electric Analogue of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics, W. P. Koenig. *AIEE CP 56-118* (available on request).

155. The Residual Time-Constant of Self-Saturating (Auto-Excited) Transductors, U. Krabbe. *Proc. IEE*, vol. 103, pt. C, Mar. 1956, pp. 71–80.

156. Diode-Shunting in Magnetic Amplifiers, J. L. Lowrance, J. D. Dolan. *AIEE Trans.*, vol. 75, pt. I, Nov. 1956, pp. 619–24.

157. Some General Properties of Nonlinear Elements, J. M. Manley, H. E. Rowe. *Proc. IRE*, vol. 44, July 1956, p. 904.

158. A. C. Controlled Transductors, A. G. Milnes, T. S. Law. *Proc. IEE*, vol. 103, pt. C, Mar. 1956, pp. 84–94.

159. A New Approach to Self-Saturating Magnetic Amplifiers, T. J. Paula, F. G. Timmel. *Proc. NEC*, vol. 12, 1956, p. 425.

160. Investigations of Magnetic Amplifiers with Feedback. Remington Rand Univac. *PB 123403*, Publications Board Project, Library of Congress, Washington 25, D. C. Mar. 1956, 20 pp.

161. Simple Reactor Circuit Operation and Mode Transition, F. J. Ringelman, A. L. Fenaroli. *Elec. Engg.*, vol. 75, Sept. 1956, p. 817.

162. Partial Analysis of the Effects of Subharmonics in the AC Supply Voltage of a Magnetic Modulator, Roy W. Roberts, Jr. *PB123404*, Publications Board Project, Library of Congress, Washington 25, D. C., Feb. 1956, 12 pp.

163. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Jalal T. Salihi. *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 296–305.

164. Transient Controlled Magnetic Amplifiers, G. Schohan. *Electronics*, Aug. 1956, p. 176.

165. The Polyunit Saturable Reactor, Karl I. Selin. *AIEE Trans.*, vol. 75, pt. III, Oct. 1956, pp. 863–67.

166. Experimental Characteristics of the 3-Phase Polyunit Saturable Reactor, Karl I. Selin, Alexander Kusko. *Ibid.*, pp. 868–71.

167. Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an Electrical Machine, H. F. Storm. *AIEE CP 56-730* (available on request).

168. Thesis Projects in Magnetic Amplifiers, H. F. Storm. *Elec. Engg.*, Oct. 1956, pp. 942–43.

169. Fast Response Multi-Stage Magnetic Amplifiers, C. C. Voice. *Proc. NEC*, vol. 12, 1956, p. 411.

170. Design Methods for Magnetic Amplifiers and Saturable Reactors, J. R. Walker, M. Frank. *OTS, PB121765*, U. S. Dept. of Commerce, July 1956, 628 pp.

171. Circuits Containing Iron-Cored Inductors and Basic Studies in Nonlinear Circuits, E. Weber. *Quarterly Report*, Polytechnic Inst. of Brooklyn, Aug.–Oct., 1956. 4 p.

172. Full-Wave Bridge Magnetic Amplifiers with Inductive Loads, H. H. Woodson. *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 7–9.

B. Magnetic Regulation and Servo Applications

1951

173. The Application of a Magnetic Amplifier to a 60-Cycle Indicating Servo System, P. W. Barnhart. *USNOL Report no. 1892*, 1951.

174. New Use for Temperature Recorders: Ready Made Amplifiers for Transducer Signals, T. R. Breunich. *Prod. Engg.*, vol. 22, Sept. 1951, pp. 143–150.

175. Design of Saturable Chokes for the Control and Regulation or Alternating Currents and Power Frequencies, H. S. Double. *Proc. IEE*, vol. 98, pt. II, no. 61, 1951, pp. 35–45.

176. Magnetic Amplifier Control for Packaged Drive. *Elec. Mfg.*, vol. 47, no. 1, 1951, p. 115.

177. Magnetic Amplifier Control for Rolling Mills. *Iron and Steel Engr.*, vol. 28, no. 11, 1951, p. 112.

178. The Control of a 65 Watt Two-Phase Servomotor with a Magnetic Amplifier, F. S. Malick. *Proc. NEC*, vol. 7, 1951, pp. 241–46.

179. A-C Arc Welders with Saturable Reactor Control, S. Oestreicher. *AIEE Trans.*, vol. 70, pt. I, 1951, pp. 787–90.

180. Prospects for Magnetic Amplifiers. *Prod. Engg.*, vol. 22, no. 5, 1951, pp. 206–08.

181. Servo Amplifier Available for Industrial Use. *Ibid.*, p. 215; also *J. Franklin Inst.*, vol. 251, June 1951, pp. 676–77.

182. The Application of a 60-Cycle Magnetic Amplifier to a Position Indicating Servomechanism, A. E. Schmid. *Proc. NEC*, vol. 7, 1951, pp. 515–22.

183. Magnetic Current Stabilizer, H. S. Sommers. *R. Sci. Inst.*, vol. 22, Aug. 1951, pp. 612–18.

184. Static Magnetic Exciter for Synchronous Alternators, H. F. Storm. *AIEE Trans.*, vol. 70, pt. I, 1951, pp. 1014–17; also *Elec. Engg.*, 1951, pp. 1084–88.

185. Voltage Regulator with Magnetic Amplifiers for Large Alternators, H. F. Storm. *Proc. NEC*, vol. 7, 1951, pp. 247–53.

186. Magnetic Amplifiers a Rising Star in

I: English Language Articles

B. Magnetic Regulation and Servo Applications

1951 (cont'd.)

Naval Electronics, A. Vincent. *J. Amer. Soc. Nav. Engrs.*, vol. 63, no. 4, Nov. 1951, pp. 903–07.

187. Study of Fundamentals and Characteristics of Magnetic Amplifiers for Flight Control Systems, J. R. Walker, D. G. Wilson. *U.S. Air Force Tech. Report no. 5881*, pt. I, Design and Analysis of Basic Circuits, Dec. 1951.

1952

188. New Magnetic Amplifier System Regulates Synchronous Machines, J. E. Barkle. *El. World*, vol. 138, no. 4, 1952, pp. 104–05; also *Westinghouse Engr.*, vol. 12, no. 6, p. 204.

189. Improvements Extend Magnetic Amplifier Applications, F. Benjamin. *Electronics*, vol. 25, no. 6, 1952, pp. 119–23.

190. The Transductor in Railway Traction Service, L. Borg, U. Krabbe. *Bull. Int. Rly. Cong. Ass.*, Elect Tract. Rlwys., vol. 3, June 1952, pp. 119–34.

191. Transrecter Auto Chargers Providing Close Float Voltage Limits, H. S. Double. *Engg. Bull.* (Siemens) no. 263–64, Feb.–May, 1952.

192. Magnetic Amplifier Voltage Regulator. *Electronics*, vol. 25, no. 1, 1952, pp. 140–42.

193. Toroidal Coil Developments, S. Freedman. *Radio-Electronic Engg.*, vol. 31, Apr. 1952, pp. 3–6.

194. A New Type of Magnetic Servo Amplifier, W. A. Geyger. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 272–80.

195. Power Control with Magnetic Amplifiers, E. L. Harder. *Electronics*, vol. 25, no. 10, 1952, pp. 115–17.

196. Magnetic Amplifier Regulation for Steel Mill Services, W. Harris, L. F. Stringer. *Blast F. & Steel Pl.*, vol. 40, no. 1, 1952, pp. 77–81.

197. Regulating Systems for the Steel Industry, W. R. Harris. *Westinghouse Engr.*, Nov. 1952.

198. Industrial Applications of Magnetic Amplifiers, W. F. Horton. *Rad. & TV News*, vol. 48, no. 11, 1952, pp. 10–12.

199. Magnetic Amplifiers for Industry, W. F. Horton. *Radio-Electronic Engg.*, Nov. 1952, pp. 10–12.

200. Development of a Modern Amplidyne-Voltage Regulator for Large Turbine Generators, W. A. Hunter, M. Temoshok. *AIEE Trans.*, vol. 71, pt. III, 1952, pp. 894–901.

201. 13 Ways to Use Magnetic Amplifiers, C. D. James. *Prod. Engg.*, vol. 23, no. 9, Sept. 1952, pp. 144–48.

202. Automatic Voltage Control, R. E. Jennings. *Elec. Times*, vol. 122, Nov. 20, 1952, pp. 911–17.

203. Performance of a New Magnetic-Amplifier-Type Voltage Regulator for Large Hydro-electric Generators, G. K. Kallenbach. *AIEE Trans.*, vol. 71, pt. II, 1952, pp. 201–05.

204. A Study of Amplifier Saturation and Magnetic Saturation in a Servo-Mechanism, C. Leondes. *AIEE CP 52-198* (not available).

205. An Improved Magnetic-Servo Amplifier, C. W. Lufcy, P. W. Barnhart, A. E. Schmid. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 281–89; also *Electronics*, vol. 25, no. 8, 1952, pp. 124–25.

206. Control of a 65 Watt 2-Phase Servo-Motor with a Magnetic Amplifier, F. S. Malick. *Elec. Engg.*, vol. 71, May 1952, pp. 466–68.

207. Saturable Reactors as R-F Tuning Elements, E. Newhall, P. Gomard, A. Ainlay. *Electronics*, vol. 25, Sept. 1952, pp. 112–115.

208. Lead Networks Utilizing a Saturable-Core Memory, D. G. Scorgie. *AIEE Trans.*, vol. 71, pt. II, 1952, pp. 297–302.

209. Automatic Regulation of Metallic Rectifiers by Magnetic Control, D. H. Smith. *Ibid.*, pt. I, 1952, pp. 111–14.

210. Performance of a New Magnetic Amplifier Type Voltage Regulator for Large Hydro-Electric Generators, H. F. Storm. *Ibid.*, pt. II, 1952, pp. 201–05.

211. A Static Magnetic Exciter, H. F. Storm. *Power Engineering*, Feb. 1952, p. 62.

212. The Transbooster; An Improved Method of Rectifier Voltage Control, A. H. B. Walker. *Electronic Engg.*, no. 12, 1952, pp. 550–56.

213. Compensation of a Magnetic Amplifier Servo System, H. H. Woodson, A. E. Schmid, C. V. Thrower. *Proc. NEC*, vol. 8, 1952, pp. 158–65.

1953

214. Steering Servo Uses Magnetic Amplifiers, J. L. Behr. *Elec. Mfg.*, May 1953.

215. Applying Magnetic Amplifiers, L. W. Buechler. *Proc. NEC*, vol. 9, 1953, p. 93.

216. A New Regulator and Excitation System, J. T. Carleton, P. O. Bobo, W. F. Horton. *AIEE Trans.*, vol. 72, pt. III, 1953, pp. 175–81.

217. Magnetic Amplifier Temperature Controller, D. Feldman. *Radio-Electronic Engg.*, June 1953, pp. 10–11.

218. Paper Mill Controls, M. H. Fisher. *Westinghouse Engr.*, Mar. 1953.

219. Industrial Regulating System, M. H. Fisher. *Ibid.*, Sept. 1953, pp. 162–67.

220. Application of a Three-Phase Motor to Servo-mechanisms, J. E. Hart. *USNRL Rept. 4168*, 1953.

221. Regulators for Marine Industry, S. A. Haverstick. *Westinghouse Engr.*, July 1953.

222. Magnetic Amplifiers, S. R. Hoh. *Tele-Tech*, vol. 12, no. 5, 1953, pp. 68–70; no. 6, pp. 114–115.

223. Magnetic Control Amplifiers XM-16A and XM-17A for Use with Servo Motors, Mark 7, Mark 8, Mark 14, and Mark 16, E. T. Hooper. *USNRL Report 2833*, 1953.

224. Magnetic Amplifiers, V. J. Louden. *GE Rev.*, vol. 26, no. 2, Mar. 1953, pp. 22–24.

225. Magnetic Amplifiers for Airborne Application, J. McKendrey. *Tele-Tech.*, July 1953, p. 35.

226. A Saturable-Core Reference Source for Use with Magnetic Amplifiers, A. G. Milnes, T. V. Vernon. *J. Sci. Inst.*, vol. 30, 1953, pp. 135–38.

227. A Range of 400 Cps and 1600 Cps Transductors for Service Use, A. G. Milnes, C. S. Hudson. *Electronic Engg.*, vol. 25, no. 306, Aug. 1953, pp. 322–26.

228. Magnetic Amplifiers in Control Circuits R. W. Moore. *Iron and Steel Engr.*, vol. 30, no. 7, July 1953, pp. 67–73.

229. Industrial Applications of Transductors. R. J. Radus. *Proc. NEC*, vol. 9, 1953, pp. 72–77.

230. Magnetic Amplifier Circuits and Applications, R. A. Ramey. *Elec. Engg.*, vol. 72, no. 9, 1953, pp. 791–95.

231. Magnetic Amplifiers, R. W. Roberts. *Machine Design*, vol. 25, Apr. 1953, pp. 171–79.

232. The Magnetic-Amplifier-Controlled Log-Carriage Drive and the Basic Duty Cycle of Carriage Performance, H. A. Rose. *AIEE Trans.*, vol. 72, pt. II, 1953, pp. 305–14.

233. What Magnetic Amplifiers Can Do to Increase Aircraft Reliability, G. F. Schroeder. *Aeron. Engg. Review*, vol. 12, no. 5, May 1953, pp. 92–95.

234. A Static Exciter for Aircraft AC Generators, D. G. Scorgie. *USNRL Rept. 4251*, 1953.

235 Applications of High Frequency Saturable Reactors, C. G. Southeimer. *Proc. NEC*, vol. 9, 1953, pp. 299–309.

236. A Magnetic Amplifier for Synchros, R. L. VanAllen. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 749–57.

237. A Magnetic Amplifier for Temperature Detection and Control, R. I. VanNice. *Proc. NEC*, vol. 9, 1953, pp. 88–92.

238. Design of Magnetic Control Amplifier XM-13A, H. H. Woodson. *OTS, PB-120929*, Mar. 1953.

1954

239. An Even-Harmonic Magnetic Amplifier and Some Applications to Measurement and Control, P. D. Atkinson, A. V. Hemingway. *Electronic Engg.*, vol. 26, no. 11, 1954, pp. 482–85.

240. Magamps Applied to Aircraft Control Problems, F. A. Baker, F. G. Timmel. *Aeron. Engg. Rev.*, no. 11, Nov. 1954, pp. 46–49.

241. Adjustable-Voltage Drives with Magnetic Amplifier Control, G. A. Bauman, C. C. Gould. *Proc. NEC*, vol. 10, 1954; also *Elec. Mfg.*, Aug. 1954, pp. 106–13.

242. Characteristics of Magnetic Amplifiers for Industrial Use, R. G. Beadle, B. P. Chausse. *Elec. Engg.*, vol. 73, no. 11, 1954, pp. 1023–27.

243. Applications of the Transductor in Rectifier Technique, L. F. Borg, P. G. Engstrom. *Direct Current*, vol. 2, no. 3, Dec. 1954, pp. 58–65.

244. Study of Voltage-Regulating Systems for Aircraft Alternating-Current Generators; Pt. I, Analogue-Computer Study and Specifications, H. C. Bourne, Jr., D. C. White. *Tech. Rept. 54-298*, Wright Air Dev. Center, Dayton, Ohio, pt. I, 1954.

245. How Good Is a Magnetic Amplifier? J. J. W. Brown. *GE Rev.*, vol. 57, no. 4, 1954, pp. 53–57; also *J. Am. Soc. Naval Engrs.*, vol. 66, no. 4, 1954.

246. Minimum Excitation Limit for Magnetic Amplifier Regulating System, J. T. Carleton, D. A. Burt, P. O. Bobo. *AIEE Trans.*, vol. 73, pt. III, Aug. 1954, pp. 869–74.

247. Magnetic Amplifiers in DC Conversion Stations, E. J. Cham, W. A. Derr. *Elec. Engg.*, vol. 73, no. 2, 1954, p. 148.

248. A 400 Cps Magnetic Amplifier for AC Servo Application, R. O. Decker, W. F. Horton. *Proc. NEC*, vol. 10, 1954, pp. 33–39.

249. Saturating Transformer Reference, W. G. Evans. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 610–16.

250. Magamp Regulator Tests and Operating Experience on West Penn Power System, E. W. Hand, F. N. McClure, P. O. Bobo, J. T. Carleton. *Ibid.*, pt. III, 1954, pp. 486–90.

251. Industrial Applications of Magnetic Amplifiers, C. G. Helmick. *Instrum. & Automation*, vol. 27, no. 8, Aug. 1954, pp. 1300–02.

252. The Transductor and Its Application, J. A. Hunt, R. H. Kelsall. *Metropolitan-Vickers Gaz.*, vol. 25, Apr. 1954, pp. 228–35.

253. Voltage Stabilization, R. E. Jennings. *Elec. Rev.*, vol. 155, no. 9, Aug. 27, 1954, pp. 315–16.

254. A New Magnetic Servo Amplifier, J. W. Kallander. *USNRL Report 4373*, 1954.

255. Transductor and Its Application, J. A. Lunt, R. H. Kelsall. *Metropolitan-Vickers Gaz.*, vol. 25, no. 417, 1954, pp. 228–35.

256. Static Excitation Control for Diesel-Electric Locomotives, S. W. McElhenny, R. M. Smith. *AIEE Trans.*, vol. 73, pt. II, 1954, pp. 158–60.

257. Magnetic Amplifier Control of Radio Frequency Generators, G. R. Mohr, R. Lee. *Ibid.*, pp. 374–78.

258. Magnetic Amplifiers and Their Application in the Mining Industry, R. D. Pettit, J. A. Heanen. *Mining Elect. Mech. Engr.*, vol. 34, May 1954, pp. 391–409.

259. An Application of Magnetic Amplifiers to Routine Testing of Junction Circuits, M. C. Stone. *Post Office El. Engrs. J.*, vol. 47, no. 1, Apr. 1954, pp. 39–40.

260. A Magnetic Amplifier for Synchros, R. L. Van Allen. *AIEE Trans.*, vol. 72, pt. I, 1953(Jan. 1954 section), pp. 749–57.

261. A Magnetic Amplifier for Temperature Detection and Control, R. I. VanNice. *Proc. NEC*, vol. 9, 1954, pp. 88–97.

262. Magnetic Amplifier Provides Automatic Gauge Control on Cold Rolling Mill, E. V. Weir. *Ibid.*, vol. 10, pp. 50–57.

263. Application of Magnetic Amplifiers, R.

J. Wessner. *Iron and Steel Engr.*, vol. 31, no. 10, 1954, pp. 101–06.

264. Lead-Lag and Lead-Integral Servo Compensation Using Half-Wave Bridge Magnetic Amplifiers, **H. H. Woodson, L. S. Weinstein, J. E. Roberts.** *USNOL Report 3560*, 1954.

265. Magnetic Amplifiers for Synchronous Motors, **M. B. Zucchino.** *Electronics*, vol. 27, pt. I, no. 3, 1954, pp. 133–35.

1955

266. Magnetic Amplifiers for Frequency Control, **S. G. Aston, M. P. Johnson.** *Brit. Commun. & Electronics*, vol. 2, no. 7, July 1955, pp. 46–49.

267. A Magnetic Amplifier Time Delay Servo, **J. L. Behr.** *Proc. NEC*, vol. 11, 1955.

268. Industrial Magnetic Amplifier Systems, **S. J. Campbell.** *Westinghouse Engr.*, May 1955, pp. 99–105.

269. 100 MVA Reactor Controlled by Direct Current, **F. J. Fisher, E. Friedlander.** *Engineering* (London), vol. 179, May 20, 1955, pp. 93–104; also *GEC J*, vol. 22, no. 2, Apr. 1955, pp. 93–104.

270. Decicycle Magnetic Amplifier Systems for Servo Applications, **L. J. Johnson, S. E. Rauch.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 667–72.

271. Magamp Regulating System for Synchronous Machines, **Kresser.** *AIEE CP 55-619* (available on request).

272. Magnetic Amplifier Control of D-C Motors, **A. Kusko, J. G. Nelson.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 326–33.

273. A Transductor-Type Field Ripple Detector for Synchronous Generators, **H. M. McConnell.** *Ibid.*, pt. III, pp. 476–79.

274. Phase-Locking of Switching-Transistor Converters for Polyphase Power Supplies, **A. G. Milnes.** *Ibid.*, pt. I, pp. 584–92.

275. 160,000 Ampere High Speed Magnetic Amplifier Design, **A. B. Rosenstein.** *Ibid.*, pt. I, pp. 90–97.

276. Speed Regulators for D.C. Motors, II, **W. Sackman.** *Elec. Mfg.*, Jan. 1955, pp. 77–82.

277. A 60-Cycle Magnetic Servo Amplifier with Electronic Preamplifier, **Schohan.** *USNOL Rept. 3976*, 1955.

278. Magnetic Servo Amplifiers Employing Tripled-Frequency Carriers, **Schohan, Ruchard.** *USNOL Rept. 3856*, 1955.

279. Regulated Power Supplies with Silicon Junction Reference, **D. G. Scorgie.** *USNRL Rept. 4589*, 1955.

280. An Ideal Component for Control Systems, **R. L. Snyder.** *Research & Engg.*, July–Aug. 1955, pp. 33–36.

281. A Magnetic Amplifier Two-Speed Servo System, **J. J. Suozzi.** *USNOL Rept. 3973*, 1955.

282. Mag-Amp. Potentiometer Servo System with Vacuum Tube or Transistor Preamp, **Wright, Raskin.** *USNOL Rept. 4024*, 1955.

283. Magnetic Amplifier Regulators in Adjustable-Speed Drives, **Wulfken.** *Elec. Mfg.*, May 1955, pp. 100–05.

284. Autopilot Magnetic Servo Amplifier, **M. Zucchino.** *Electronics*, vol. 28, pt. I, May 1955, pp. 208–14.

1956

285. A Line Voltage Regulator Having Magnetic Amplifier Control, **F. W. Anderson.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 108–12.

286. A Magnetic Amplifier Grid Control Circuit, **J. H. Burnett.** *Proc. IRE*, Apr. 1956, p. 529.

287. Magnetic Amplifiers for High-Performance Power Servos, **G. P. deWestfelt.** *Elec. Mfg.*, Nov. 1956, p. 115.

288. Applying Magnetic Amplifiers, **W. F. Eagen.** *Allis-Chalmers Elec. Rev.*, vol. 21, no. 3, 1956, pp. 20–29.

289. A High Temperature Regulated D.C. Power Supply for Aircraft, **R. G. Engman, R. E. King.** *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 170.

290. A Seventy-Five KW Magnetic Amplifier, **K. Enslein.** *Ibid.*, p. 161.

291. A Voltage Regulator Using High Speed of Response Magnetic Amplifiers With Transistor Driver, **E. Fthenakis.** *Ibid.*, p. 185.

292. Measurements on High Speed Magnetic Servo Amplifiers With Two Phase Motor Load, **W. A. Geyger.** *Elec. Engg.*, vol. 75, June 1956, p. 518.

293. Self-Balancing Magnetic Servo Amplifier, **W. A. Geyger.** *Electronics*, vol. 29, Mar. 1956, pp. 196–99.

294. Magnetic Amplifier Controlled Regulated Rectifiers, **H. L. Goldstein, R. J. Lowell.** *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 145.

295. Inverse-Time Under-Frequency Relay, **R. G. Hoft.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 625–29.

296. The Design of a Completely Static Accurately Regulated Power Supply, **R. G. Hoft, N. A. Read.** *AIEE CP 56-959* (available on request).

297. Magnetic Amplifiers in Heavy-Duty Systems, **U. Krabbe.** *Prod. Engg.*, vol. 27, no. 1, 1956, p. 2.

298. A Magnetic Voltage Reference, **E. W. Manteuffel, R. O. McCrary.** *Proc. STC on Mag. Amp., T*-86, Apr. 1956, p. 200.

299. Comparison of Rotating, Electronic, and Magnetic Amplifier Regulators, **J. P. Montgomery.** *Elec. Engg.*, vol. 75, Apr. 1956, p. 359.

300. Magnetic Amplifiers on a High Speed Accuracy Single Stand Reversing Cold Mill, **A. L. Mozina.** *Blast F & Steel Pl.*, June 1956, p. 636.

301. Magnetic Amplifiers for Control of Servo-motors, **A. R. Perrins.** *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 246.

302. Regulated Power Supplies with Silicon Junction Reference, **D. G. Scorgie.** *Ibid.*, p. 150.

303. Magnetic Amplifier Two Speed Servo System, **J. J. Suozzi.** *Electronics*, vol. 29, Feb. 1956, p. 140.

C. Magnetic Frequency Conversion

1951

304. The Magnetic Cross Valve, **H. J. McCreary.** *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 1868–75.

1952

305. Subharmonic Oscillations in Electric Circuits Containing Iron-Core Reactors, **J. P. Schouten, H. J. Heijn.** *Appl. Sci. Research*, (Hague), pt. B, vol. 2, no. 4, 1952, pp. 301–19.

1953

306. The Frequency Tripler, **L. R. Blake.** *Proc. IEE*, vol. 100, pt. II, 1953, pp. 296–309.

307. Preliminary Development of a Magnettor Current Standard, **E. P. Felch, J. L. Potter.** *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 524–31.

308. Magnetic Frequency Conversion, **L. C. Harriott.** *Proc. NEC*, vol. 9, 1953, pp. 78–87.

309. Magnetic Converter D-C Amplifier, **W. A. Rote.** *Electronics*, vol. 26, no. 12, 1953, pp. 170–73.

1954

310. Magnetic Frequency Multipliers, **L. J. Johnson, S. E. Rauch.** *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 448–51.

1955

311. Saturable Reactor Frequency Divider, **G. W. C. Court, C. I. C. Scollay.** *Wireless Engr.*, vol. 32, no. 12, Dec. 1955, pp. 328–29.

312. Odd Integer Magnetic Frequency Multipliers, **L. J. Johnson, S. E. Rauch.** *Proc. IRE*, vol. 43, no. 2, Feb. 1955, pp. 168–73.

313. Decicycle Magnetic Amplifier Systems for Servo Applications, **L. J. Johnson, S. E. Rauch.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 667–72.

314. Economic Design of Saturating Reactor Magnetic Pulsers, **R. A. Mathias, E. M. Williams.** *Ibid.*, pp. 169–71.

315. Magnetic Servo Amplifiers Employing Tripled-Frequency Carriers, **G. Schohan, Ruchard.** *USNOL Rept. 3856*, 1955.

316. Analysis and Design of a Magnetic Frequency Multiplier, **O. J. M. Smith, J. T. Salihi.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 99–106.

317. An All Magnetic Audio Amplifier System, **J. J. Suozzi, E. T. Hooper.** *Ibid.*, pp. 297–300.

318. A New Self-Excited Square-Wave Transistor Power Oscillator, **G. C. Uchrin, W. O. Taylor.** *Proc. IRE*, vol. 43, no. 1, 1955, p. 99.

1956

319. Magnetic Arbitrary Waveform Generator, **C. B. House.** *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 235.

320. Magnetic Frequency Multipliers and Their Rating: Pt. I, Frequency Triplers, **W. McMurray.** *AIEE Trans.*, vol. 75, pt. I, pp. 384–90.

321. Partial Analysis of the Effects of Subharmonics in the AC Supply Voltage of a Magnetic Modulator, **Roy W. Roberts, Jr.** *PB 123404*, Publications Board Project, Library of Congress, Washington 25, D. C., Feb. 1956, 12 pp.

322. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Materials in Series with Capacitors, **J. T. Salihi.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 296–305.

323. Static Frequency Multipliers for Induction Motors, **G. Schohan.** *Elec. Mfg.*, vol. 57, Apr. 1956, p. 136.

D. Operational and Instrument Magnetic Amplifiers

1951

324. The Magnetic Modulator, **R. Feinberg.** *Wireless Engr.*, vol. 28, Sept. 1951, pp. 281–88.

325. Magnetic Amplifiers of the Balance-Detector Type—Their Basic Principles, Characteristics and Applications, **W. A. Geyger.** *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 1707–20.

326. The Inductor Compass, **A. Hine.** *Proc. IEE*, vol. 98, pt. II, no. 64, 1951, pp. 485–98.

327. Isolation Metering of D. C. Bus Current, **W. F. Horton.** *Proc. NEC*, vol. 7, 1951, pp. 260–62.

328. A Simple Magnetic Modulator for Conversion of Millivolt DC Signals, **G. Wennerberg.** *Elec. Engg.*, vol. 70, Feb. 1951, pp. 144–47.

1952

329. Magnetic Modulators, **E. P. Felch, V. E. Legg, F. G. Merrill.** *Electronics*, Feb. 1952, pp. 113–17.

330. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, **W. A. Geyger.** *AIEE Trans.*, vol. 71, 1952, pp. 383–95.

331. Magnetic Converter Amplifier. *Instruments*, vol. 25, no. 4, 1952.

332. The Single Core Magnetic Amplifier as a Computer Element, **R. A. Ramey.** *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 442–46.

1953

333. Magnetic-Amplifier Applications in D-C Conversion Stations, **E. J. Cham, W. A. Derr.** *AIEE Trans.*, vol. 72, pt. III, 1953, pp. 220–28.

334. The Magnetic Amplifier as an Analogue Computer Component, **L. J. Craig.** *Proc. IEE*, vol. 41, Oct. 1953, pp. 1477–82.

335. Preliminary Development of a Magnettor Current Standard, **E. P. Felch, J. L. Potter.** *AIEE Trans.*, vol. 72, pt. I, Nov. 1953, pp. 524–31.

I: English Language Articles

D. Operational and Instrument Magnetic Amplifiers

1953 (cont'd.)

336. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, **L. A. Finzi, R. A. Mathias.** *Ibid.*, Sept., pp. 455–61.

337. The Study of an Inverter for Amplification for Low Input Power DC Signals, **E. H. Frost-Smith.** *Proc. IEE*, vol. 100, pt. II, Aug. 1953, pp. 362–74.

338. The New AEG Vectormeter, **F. Koppelmann.** *AEG Prog.*, New York, N. Y., no. 3, July 1953, pp. 77–81.

339. Electrons at Work; Magnetic Amplifier Phase Shift Corrector, **A. A. McKenzie.** *Electronics*, vol. 26, no. 12, Dec. 1953, 208–09.

340. The Magnetic Modulator in A-C Servo Corrective Networks, **C. Volz.** *Proc. NEC*, vol. 9, Sept. 1953, pp. 209–14.

1954

341. Magnetic Demodulator for Use in Magnetic Amplifiers, **J. L. Behr.** *Proc. NEC*, vol. 10, 1954, pp. 25–32.

342. Transductors for Instrumentation of D-C Circuits, **E. J. Cham. W. A. Derr.** *Westinghouse Engr.*, vol. 14, no. 2, 1954, pp. 86–89.

343. Transductors as D-C Instrument Transformers, **E. J. Cham, W. A. Derr.** *Elec. Mfg.*, Mar. 1954, pp. 136–41.

344. An Analogue Computer Technique Using Magnetic Amplifiers, **B. E. Davis, I. H. Swift.** *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 635–40.

345. Magnetic Amplifiers in Metering Direct Current on Electrolytic Cell Lines, **E. A. Downing.** *Ibid.*, pt. II, no. 12, 1954, pp. 93–96.

346. Composite Cores for Instrument Transductors, **E. H. Frost-Smith, A. E. DeBarr.** *Proc. IEE, Paper no. 1636*, 1954.

347. An Operational Type Magnetic Amplifier for Air-Borne Servo Control Systems, **R. M. Hubbard.** *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 425–30.

348. Three Phase High Speed Magnetic Amplifier, **A. E. Maine.** *Electronic Engg.*, vol. 26, no. 315, 1954, pp. 180–85.

349. Pulse Relaxation Amplifier—A Low-Level DC Magnetic Amplifier, **R. E. Morgan, J. B. McFerran.** *AIEE Trans.*, vol. 73, pt. I, July 1954, pp. 245–49.

350. Industrial Applications of Transductors, **R. J. Radus.** *Tele-Tech & Elec. Ind.*, vol. 13, no. 4, 1954, pp. 78–79.

1955

351. Transductors for Measurement, **L. Borg.** *ASEA J.*, vol. 28, no. 7, July 1955, pp. 110–18.

352. Operational Magnetic Amplifier with Audio Frequency Transistor Power Supply, **R. O. Decker, F. Gourash.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 490–96.

353. Low Level Magnetic Amplifier, **F. Gourash.** *Tele-Tech & Elec. Ind.*, vol. 14, July–Dec. 1955, pp. 90–91.

354. The Application of Perpendicularly Superposed Magnetic Fields, **R. A. Heartz, H. Buelteman, Jr.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 655–59.

355. Magnetic Amplifiers in Fire Control Computers, **E. D. Hooper.** *USNOL Report no. 4034*, 1955.

356. Zero Signal Stability and Noise Suppression with Magnetic Demodulator-Amplifiers, **L. J. Johnson, S. E. Rauch.** *Proc. NEC*, vol. 11, Oct. 1955, pp. 145–51.

357. A Static Transistor Analogue Divider, **D. H. Schaefer.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 652–55.

358. Four-Quadrant Multiplication with Transistors and Magnetic Cores, **R. L. Van Allen.** *Ibid.*, pp. 643–48.

359. Saturating Reactor Magnetic Pulser Design, **E. M. Williams, R. A. Mathias.** *Elec. Engg.*, vol. 74, no. 6, June 1955, p. 525.

1956

360. USAF Computer Uses Magnetic Amplifiers, **J. M. Carroll.** *Electronics*, vol. 29, July 1956, p. 138.

361. A Stable Low-Level Magnetic Amplifier, **R. Cockrell.** *AIEE CP 56-737* (not available).

362. Transductor Applications, **W. A. Derr, E. J. Cham.** *Westinghouse Engr.*, vol. 16, no. 3, May 1956, pp. 82–84.

363. A Magnetic Amplifier Simulator, **R. E. Fitts.** *AIEE CP 56-92* (available on request).

364. Sixty Cycle Self-Balancing Magnetic Servo Amplifier, **W. A. Geyger.** *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 221.

365. Transistor Magnetic Analogue Multiplier, **G. L. Keister.** *Electronics*, vol. 29, Oct. 1956, pp. 160–63.

366. A Signal Discriminating Magnetic Amplifier, **G. E. Lynn, J. F. Ringelman.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 97–102.

367. The Pulse Stretch Coupling Circuit, **H. W. Patton.** *Ibid.*, pp. 377–79.

368. Transcendental Function Analogue Computation with Magnetic Cores, **D. H. Schaefer, R. L. Van Allen.** *Ibid.*, pp. 160–65; also *OTS, PB 111900*, Feb. 1956, 11 pp.

369. Core Volume Derivation for Magnetic Pulse Modulators, **J E. Sunderlin.** *AIEE CP 56-228* (available on request).

E. Static Magnetic Switching Applications

1951

370. The Use of Saturable Reactors as Discharge Devices for Pulse Generators, **W. S. Melville.** *Proc. IEE*, vol. 98, pt. 3, 1951, pp. 185–207.

1952

371. The Application of Transductors as Relays to Protective Gear, **R. K. Edgley, F. L. Hamilton.** *Proc. IEE*, vol. 99, pt. II, no. 70, 1952, pp. 297–313.

372. Applications of Transductors. Their Use in Protective Gear. *Elec. Rev.*, vol. 150, no. 3879, 1952.

1953

373. Interesting Nonlinear Effects, **R. S. Mackay.** *J. Applied Physics*, vol. 24, Mar. 1953, pp. 311–13.

374. Same Aspects of Transductor Design With Particular Reference to Discharge Lamp Circuits, **C. E. Williams.** *Proc. IEE*, vol. 100, pt. II, no. 78, 1953, pp. 667–73.

1954

375. Magnetic Switching Network for Data Handling Systems, **L. P. Gieseler.** *Proc. NEC*, vol. 10, 1954, pp. 156–61.

376. Testing Magnetic Decision Elements, **J. D. Goodwell.** *Electronics*, vol. 27, pt. I, Jan. 1954, pp. 200–03.

377. Magnetic Switching Circuits, **R. C. Minnick.** *J. Applied Physics*, Apr. 1954.

378. Saturable Transformers as Gates, **B. Moffat.** *Electronics*, Sept. 1954, pp. 174–78.

379. A Review of Magnetic and Ferroelectric Computing Components, **V. L. Newhouse.** *Electronic Engg.*, May 1954, pp. 192–99.

1955

380. Magnetic Amplifiers for Indication and Control, **Collins, Fields.** *Elec. Mfg.*, Sept. 1955, pp. 112–16.

381. Magnetic Amplifier Sequential Circuits of the Asymmetrical Type, **A. S. Fitzgerald.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 685–91.

382. Economic Design of Saturating Reactor Magnetic Pulsers, **R. A. Mathias, E. M. Williams.** *Ibid.*, pp. 169–71.

383. The Transfluxor: A Magnetic Gate with Stored Variable Setting, **J. A. Rajchman, A. W. Lo.** *RCA Rev.*, vol. 16, no. 2, June 1955, pp. 303–11.

1956

384. Ferroresonant Computing Circuits, **R. S. Arbon, G. Phylip-Jones.** *Wireless World*, vol. 62, July 1956, pp. 324–30.

385. Cypak Systems—An Application of Logic Functions, **J. P. Baker.** *Westinghouse Engr.*, July 1956, p. 107.

386. Basic Magnetic Switching Circuits, **R. W. Durkee.** *Elec. Mfg.*, Oct. 1956, p. 102.

387. Magnetic Logic Circuits for Industrial Control Systems, **W. G. Evans, W. G. Hall, R. I. Van Nice.** *AIEE Trans.*, vol. 75, pt. II, 1956, pp. 166–71.

388. Magnetic Trip Amplifier, **Hodson, E. W. Manteuffel.** (available from the authors)

389. A Magnetic Amplifier Switching Matrix, **D. Katz.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 236–41.

390. The Transfluxor, **J. A. Rajchman, A. W. Lo.** *Proc. IRE*, Mar. 1956, p. 321.

391. Magnetic Amplifiers for Replacement of Relays, **R. W. Roberts, R. J. Radus, O. R. Rielsy, C. V. Fields.** *Proc. NEC*, vol. 12, 1956, p. 454.

392. Magnetic Logic Circuit Control System Design Considerations, **R. I. Van Nice.** *AIEE Trans.*, vol. 75, pt. I, Nov. 1956, pp. 595–600.

F. Transistor–Magnetic Device Combinations

1953

393. Applications of Transistors to the Control of Magnetic Amplifiers, **G. F. Pittman, Jr.** *Proc. NEC*, vol. 9, 1953, pp. 64–71.

394. Transistor Controlled Magnetic Amplifier, **R. H. Spencer.** *Electronics*, vol. 26, Aug. 1953, pp. 136–40.

395. Transistor-Controlled Half-Wave Magnetic-Amplifier, **J. J. Suozzi.** *OTS, USNOL PB 120908*, July 1953.

1954

396. Transistor Control of Magnetic Amplifiers, **G. F. Pittman, Jr.** *Radio-Electronic Engg.*, Feb. 1954, pp. 13–15; also *Rad. & TV News*, vol. 5, no. 2, 1954, p. 13.

397. Transistor Preamplifier Drives Magnetic Servo, **M. B. Zucchino.** *Electronics*, vol. 27, pt. I, no. 3, 1954, pp. 168–71.

1955

398. Transistor Switch Improves Magnetic Amplifier Performance, **R. L. Bright, R. O. Decker, R. A. Ramey.** *Elec. Mfg.*, Apr. 1955, pp. 98–101.

399. The Use of Junction Transistors as Switches, **R. L. Bright.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 111–21.

400. Transistor Demodulator for High-Performance Magnetic Amplifiers in A.C. Servo Applications, **R. O. Decker.** *Ibid.*, pp. 121–23.

401. Phase-Locking of Switching-Transistor Converters for Polyphase Power Supplies, **A. G. Milnes.** *Ibid.*, pp. 584–92.

402. DC to AC Square-Wave Static Converter and Amplifier, **H. T. Mortimer.** *USNRL Report 4582*, 1955.

403. A High-Accuracy Static Time Delay Device Utilizing Transistors, **G. F. Pittman, Jr.** *AIEE Trans.*, vol. 74, pt. 1, 1955, pp. 54–58.

404. A Switching Transistor DC to AC Converter Having an Output Voltage Proportional to DC Input Voltage, **G. H. Royer.** *Ibid.*, pp. 322–26.

405. A Static Transistor Analogue Divider, D. H. Schaefer. *Ibid.*, pp. 652–55.

406. A New Self-Excited Square-Wave Transistor Power Oscillator, G. C. Uchrin, W. O. Taylor. *Proc. IRE* vol. 43, no. 1, 1955, p. 99.

407. Variable Frequency Magnetic Coupled Multivibrator, R. L. Van Allen. *AIEE Trans.*, vol. 74, pt. 1, 1955, pp. 356–61.

408. Four Quadrant Multiplication with Transistors and Magnetic Cores, R. L. Van Allen. *Ibid.*, pp. 643–48.

409. Mag.-Amp. Potentiometer Servo System with Vacuum Tube or Transistor Preamp., Wright, Raskin. *USNOL Report 4024*, 1955.

1956

410. A Single-Transistor Magnetic-Coupled Oscillator, Kan Chen, A. J. Schiewe. *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 396–99.

411. Magnetic Amplifier Control of Switching Transistors, H. W. Collins. *Ibid.*, Nov., pp. 585–89.

412. Magnetic Amplifier Control of Switching Transistors, H. W. Collins. *Elec. Engg.*, vol. 75, Sept. 1956, p. 812.

413. A Voltage Regulator Using High Speed of Response Magnetic Amplifiers with Transistor Driver, E. Fthenakis. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 185.

414. A High Current Switch for a Transistor Operated Magnetic Memory, W. A. Helbig, W. G. Rumble. *Proc. NEC*, vol. 12, 1965, p. 300.

415. Transistor-Magnetic Analog Multiplier, G. L. Keister. *Electronics*, vol. 29, no. 10, Oct. 1956, pp. 160–63.

416. High-Gain Transistor Magnetic Amplifier, E. Stuhlinger, J. C. Taylor, D. L. Anderson. *Elec. Mfg.*, vol. 57, no. 3, Mar. 1956, pp. 102–05.

417. A Transistor Operated Magnetic Core Memory, C. S. Warren, M. G. Rumble, W. A. Helbig. *Proc. NEC*, vol. 12, 1956, p. 289.

G. Magnetism and Magnetic Materials

1951

418. High Saturation Magnetic Alloy with Rectangular Hysteresis Loop, J. F. Libsch, E. Both. *Elec. Engg.*, vol. 70, no. 5, 1951, pp. 420–21.

1952

419. Some Pulse Tests on Magnetic Specimens Having Rectangular Hysteresis Loops, F. F. Dorey. *Soft Magnetic Materials in Telecommunication*, London, England, Apr. 1952, pp. 286–300.

420. Characteristics of Overlapping Joints in Magnetic Circuits, H. L. Garbarino, T. D. Gordy. *AIEE Trans.*, vol. 71, pt. III, 1952, pp. 386–92.

421. The Calculation of the Magnetizing Force, A. A. Halacsy. *Ibid.*, pt. I, pp. 90–95.

422. Survey, Characteristics and Evaluation of High Performance Magnetic Core Materials, S. R. Hoh. *OTS, No. 52-287; PB111211*, Wright Air Development Center, Nov. 1952.

423. The Effect of Core Materials on Magnetic Amplifier Circuits, L. J. Johnson. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 26–31.

424. Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis Loops, M. F. Littman. *Ibid.*, pp. 220–23.

425. The Behavior of Rectangular Hysteresis Loop Materials Under Current Pulse Excitation, E. Sands. *Proc. IRE*, vol. 40, no. 10, 1952, pp. 1246–50.

1953

426. A. Review of New Magnetic Phenomena, R. E. Alley, Jr. *BSTJ.*, vol. 32, 1953, p. 1155.

427. Evaluation of Core Materials for Magnetic Amplifiers, D. C. Dieterly. *Elec. Mfg.:* Part I, Jan. 1953, p. 68; Part II, Feb. 1953, p. 124.

428. Mathematical Descriptions of Core Losses, J. W. Hale, F. R. Richardson. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 495–501.

429. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, H. W. Lord. *Ibid.*, pp. 85–88.

430. The Influence of Magnetic-Amplifier Circuits Upon the Operating Hysteresis Loops, H. W. Lord. *Ibid.*, pp. 721–28.

431. Stressed Ferrites Having Rectangular Hysteresis Loops, H. J. Williams, R. C. Sherwood, M. Goertz, F. J. Schnettler. *Ibid.*, pp. 531–37.

1954

432. Flux Resetting Characteristics of Several Magnetic Materials, H. Huhta. *AIEE Trans.*, vol. 73, pt. I, May 1954, pp. 111–14.

433. Eddy Current Phenomena in Ferromagnetic Materials, H. M. McConnell. *Ibid.*, pp. 226–35.

434. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *Ibid.*, pp. 682–90.

1955

435. Effects of Transverse Compressional Stress on Magnetic Laminations, R. E. Fischell. *AIEE Trans.*, vol. 74, pt. I, May 1955, pp. 190–93.

436. Influence of ID/OD Ratio on Static and Dynamic Magnetic Properties of Toroidal Cores, R. W. Roberts, R. I. Van Nice. *Ibid.*, pp. 599–607.

437. Relation of D-C Magnetic Properties of Oriented 48 per cent Nickel-Iron to Magnetic Amplifier Performance, C. E. Ward, M. E. Littmann. *Ibid.*, pp. 422–27.

1956

438. The Process of Magnetization, C. P. Bean. *Proc. STC on Mag. Amp., T 86*, Apr. 1956, p. 30.

439. Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers, C. L. Boyajian. *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 307–13.

440. Dynamic Core Behavior and Magnetic Performance, L. A. Finzi, D. L. Critchlow. *Ibid.*, vol. 76, pt. I, 1957, pp. 229–40.

441. Magnetic Effects of Compressional Stress at Low Field Intensities, R. E. Fischell. *Ibid.*, vol. 75, pt. I, 1956, pp. 148–56.

442. Flux Reversal in Magnetic Amplifier Cores, F. J. Friedlaender. *Ibid.*, pp. 268–78.

443. Dynamax—A New Crystal and Domain-Oriented Magnetic Core Material, G. H. Howe. *Ibid.*, Nov., pp. 548–50; also *Elec. Engg.*, Aug. 1956, p. 702.

444. Evaluation of 6.4% Silicon-Iron Alloy for Use in Magnetic Amplifiers, E. J. O'Brien. *USNOL Report 4233*, 1956.

445. Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops, C. B. Wakeman, F. J. Beck. *AIEE Trans.*, vol. 75, pt. I, Nov. 1956, pp. 562–69.

H. Measurements and Testing

1952

446. Magnetic Amplifier Gapless-Core Tests, J. R. Conrath. *Electronics*, Nov. 1952, pp. 119–21.

447. Dynamic Hysteresis Loop Measuring Equipment, H. W. Lord. *AIEE Trans.*, vol. 71, pt. I, 1952, pp. 269–72.

1953

448. An Improved Form of Hysteresis Loop Plotter for Magnetic Materials (note), A. D. Booth. *J. Sci. Inst.*, vol. 30, Oct. 1953, pp. 384–85.

449. Nondestructive Sensing of Magnetic Cores, D. A. Buck, W. I. Frank. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 822–30.

450. Butterfly Curve Tracer for Magnetic Materials, G. M. Ettinger. *Electronics*, vol. 26, Mar. 1953, pp. 118–21.

451. Magnetic Amplifier Transient Analyzer, W. A. Geyger. *Ibid.*, no. 7, 1953, pp. 189–91.

452. An AC-DC Gauss Meter, A. Kaufmann. *Instruments*, vol. 26, 1953, p. 1360.

453. A Permeability Analyzer for Magnetic Amplifier Cores, P. Siskind. *AIEE Trans.*, vol. 72, pt. I, 1953, pp. 572–76.

454. A Transient Analyzer for Magnetic Amplifiers, E. J. Smith. *Ibid.*, pp. 461–65.

1954

455. Recording Fluxmeter, R. I. Berge, G. A. Guderjohn. *Electronics*, vol. 27, pt. 2, July 1954, pp. 147–49.

456. The Cyclic Integrator—A Device for Measuring the Frequency Response of Magnetic Amplifiers, T. Dunnegan, Jr., J. Harnden, Jr. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 358–65.

457. Testing Core Materials (Permanent Cup Method), L. J. Johnson. *Elec. Mfg.*, May 1954, pp. 130–33.

458. Techniques for Measuring Cascaded Self-Saturating Magnetic Amplifier Performance, H. Kaplan, G. Wolff. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 581–84.

459. An Automatic Plotter for Magnetic Hysteresis Loops, H. McGross. *Proc. IEE*, vol. 101, pt. II, no. 82, 1954, pp. 417–30.

460. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 682–90.

461. A Transient Analyzer for Magnetic Amplifiers, E. J. Smith. *Ibid.*, pp. 461–65.

462. Evaluation of Core Materials for Magnetic Amplifier Applications, R. D. Teasdale, H. R. Brownell. *Conv. Rec. IRE*, vol. 3, 1954, pp. 56–65.

1955

463. Standards for the Presentation of Data on Magnetic Amplifier Core Materials, AIEE Mag. Amp. Subcommittee. *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 598–99.

464. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, D. L. Critchlow. *Ibid.*, pp. 607–10.

465. An Accurate Electronic Tracer for Dynamic Characteristics of Magnetic Materials, E. Kittl. *Ibid.*, pp. 407–18.

1956

466. Some Frequency Response Measurements on Magnetic Amplifiers, R. C. Barker, G. M. Northrop. *Proc. NEC*, vol. 12, 1956, p. 444.

467. A Sensitive Single Turn Hysteresis Loop Tester, F. Bernstein, T. H. Bonn, R. D. Torrey. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 106.

468. Magnetic Switch B/H Loop Tracer, W. A. Geyger. *Electronics*, vol. 29, Oct. 1956, pp. 167–68.

469. The Ferrotracer—A Simple Dynamic B-H Tracer with Silicon Diode Chopper Circuits, W. A. Geyger. *Proc. NEC*, vol. 12, 1956, p. 83.

470. Simple 60 cps Hysteresis Loop Tracer for Magnetic Materials of High or Low Permeability, D. H. Howling. *R. Sci. Inst.*, vol. 27, Nov. 1956, pp. 952–56.

471. Design Criteria for a Practical Flux Reset Core Tester, J R. Jaquet. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 74.

472. A Delta-B Indicator, H. W. Lord. *Elec. Engg.*, vol. 75, Nov. 1956, p. 1012.

473. Correlation of Hysteresis Loop Core Tests and Magnetic Amplifier Control Characteristics, W. J. Muldoon. *Proc. STC on Mag. Amp., T-86*, Apr. 1956, p. 116.

474. Present Methods of Core Evaluation, Grading, and Matching, R. W. Roberts. *Ibid.*, p. 31.

475. Magnetic Measurements, C. E. Webb. *Metallurgia*, Manchester, England, vol. 54, Aug. 1956, pp. 57–66.

I: English Language Articles

J. Associated Devices

1952

476. Titanium Dioxide Rectifiers, **E. S. Breckenridge, W. R. Hosler.** *Journal of Research*, National Bureau of Standards, vol. 49, no. 2, Aug. 1952, *Research Paper 2344*, pp. 65–72.

477. The Use of Selenium Rectifiers in Self-Saturating Magnetic Amplifiers, **J. R. Conrath.** *Elec. Engg.*, vol. 71, no. 1, 1952, pp. 50–54.

478. Germanium Rectifiers, **E. A. Harty.** *AIEE Pub. S-50*, 1952, pp. 74–76.

1953

479. Selenium Rectifiers, Factors in Their Application, **J. Gramels.** *BSTJ*, vol. 32, Nov. 1953, pp. 1469–92.

480. The Manufacture and Testing of Germanium Rectifiers, **C. Peters.** *GEC Telecomm.*, no. 17, 1953, p. 34.

1954

481. Magnetic Amplifier Uses Conventional Inductors, **A. I. Bennett, Jr.** *Electronics*, vol. 27, pt. 1, no. 1, 1954, pp. 181–83.

482. Instantaneous Electrical Characteristics of Selenium Rectifiers, **G. F. Pittman, Jr.** *AIEE Trans.*, vol. 73, pt. 1, 1954, pp. 45–49.

483. Rectifiers in Magnetic Amplifier Circuits, **B. J Seddon.** *Proc. NEC*, vol. 10, 1954, pp. 168–81.

484. The Use of Junction Transistors as Switches, **R. L. Bright.** *AIEE Trans.*, vol. 74, pt. I, 1955, pp. 111–21.

485. A Design Method for a Silicon Diode Reference Device, **D. A. Burt.** *Proc. NEC*, vol. 11, Oct. 1955, pp. 907–18.

1956

486. Current References for Magnetic Amplifiers, **D. A. Burt.** *Electronic Ind. & Tele-Tech.*, vol. 15, no. 10, 1956, p. 66.

487. Semi-Conductor Power Rectifiers in Continuous Duty Circuits, **F. W. Gutzwiller.** *Elec. Mfg.*, July 1956, p. 109.

488. High-Temperature Selenium Rectifiers—A Survey of Manufacturers' Data, **S. Shilliday.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 248–51.

K. Ferroresonance and Its Applications

1952

489. Ferroresonant Flip-Flops, **C. Isborn.** *Electronics*, vol. 25, Apr. 1952, pp. 121–23.

1954

490. Forced Oscillations of Non-Linear Circuits, **L. A. Pipes.** *AIEE Trans.*, vol. 73, pt. I, 1954, pp. 352–58.

491. Ferroresonant Flip-Flop Design, **R. W. Rutishauser.** *Electronics*, vol. 27, May 1954, pp. 152–53.

1956

492. Ferroresonant Computing Circuits, **R. S. Arbon, G. Phylip-Jones.** *Wireless World*, vol. 62, July 1956, pp. 324–30.

493. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, **J. T. Salihi.** *AIEE Trans.*, vol. 75, pt. I, 1956, pp. 296–307.

494. Transient Behavior in a Ferroresonant Circuit, **J. G. Skalnik.** *J. Applied Physics*, vol. 27, May 1956, pp. 508–13.

L. Manufacturing Techniques, Environmental Effects, and Reliability

1953

495. Short Time Thermal and Mechanical Limits for Transformers and Reactors, **J. E. Clem.** *AIEE Trans.*, vol. 72, pt. III, June 1953, pp. 432–47.

1954

496. Miniature Amplistat for Aircraft Applications at 400° F Ambient Temperature, **R. H. Ball.** *Proc. NEC*, vol. 10, 1954, pp. 40–49.

1955

497. Etched Wiring Simplifies Magnetic Amplifier Design, **Sueker.** *Elec. Mfg.*, Jan. 1955, p. 106.

1956

498. Impregnation of Toroids for High Temperature Service, **E. O. Demell.** *Transactions*, Professional Group on Component Parts, IRE, vol. CP-3, no. 3, Dec. 1956, p. 113.

499. Quick Wind Device for Toroidal Coil Samples, **R. A. MacMillan.** *R. Sci. Inst.*, vol. 27, no. 7, July 1956.

II: Foreign Language Articles

A. Theory, Analysis, and Design

1951

500. Graphical-Analytical Method for Calculating Transient Response in Electrical Circuits Containing Nonlinear Elements, **L. A. Bessonov.** *Electricity* (USSR), no. 5, 1951, pp. 66–70.

501. The Magnetic Amplifier, **H. Feigs.** *Funk und Ton* (Berlin), vol. 5, no. 4, 1951, pp. 210–20.

502. The Magnetic Amplifier, **R. Feinberg.** *Bull. de l'Association Suisse Electr.* (Zurich), vol. 42, no. 5, 1951, p. 148.

503. The Time Constant of Magnetic Amplifiers, **H. Goldstein.** *Funk und Ton* (Berlin), vol. 5, no. 2, 1951, pp. 76–78.

504. Simplified Saturable Reactor Design, **D. I. Maryanovsky, I. A. Kazantseva.** *Electricity* (USSR), no. 3, 1951, pp. 40–47.

505. Curves for the Simultaneous Magnetization of a Feedback Amplifier by AC and DC Fields, **I. B. Negnevitsky.** *Ibid.*, no. 10, pp. 62–72.

506. A Ferro Magnetic Amplifier, **A. Rauth.** *Électricien* (Paris), vol. 79, no. 1898, 1951, pp. 195–200.

507. Fundamentals of Magnetic Amplifier Theory. Part II: Firing Angle Control, **W. Schilling.** *ETZ* (Berlin), vol. 72, no. 15, Ausg. A, 1951, pp. 465–69.

1952

508. Design of Low Power Magnetic Bridge Amplifiers, **V. L. Benin.** *Collected Works*, Institute of Electr. Eng. of the Acad. Sci. Ukranian SSR, no. 9, 1952, pp. 71–83.

509. Transient Response in Multi-Stage Magnetic Amplifiers with Internal Feedback, **L. A. Grigoryan.** *Proc. Acad. Sci. Armenian SSR*, Erevan, Armenia, vol. 15, no. 1, 1952, pp. 17–21; no. 3, pp. 79–84; no. 5, pp. 135–38.

510. Magnetic Amplifiers, **M. Gurevich.** *Radio-Engineering* (USSR), no. 9, 1952.

511. Magnetic Amplifiers, **W. Kramer.** *ETZ* (Berlin), Iss. 7, Ausg. A, Apr. 1952, pp. 210–11.

512. The Theory of Magnetic Amplifiers, **F. Ludi.** *Z. angew. Math. Phys.* (Basle), vol. 3, no. 2, 1952, pp. 119–28.

513. Setting the Control with Magnetic Amplifiers, **O. Mohr.** *ETZ* (Berlin), Iss. 7, Apr. 1952, pp. 205–06.

514. The Transductor and the Transistor, **J. Moline.** *Radio Francaise* (Paris), no. 4, 1952, pp. 1–5.

515. Approximate Design of Magnetic Amplifiers, **I. D. Pashentsev.** *Collected Sci. Works*, Leningrad Power Inst. of Railroad Eng., no. IV, 1952, pp. 98–123.

516. Magnetic Amplifiers, **B. Pistoulet.** *Rev. Gen. de l'Électricité* (Paris), no. 1, 1952, pp. 45–52.

517. Properties and Applications of Magnetic Amplifiers, **V. Pollack.** *Slaboproudy Obzor* (Prague), vol. 13, 1952.

518. The Magnetic Amplifier, **J. J. Reitveld.** *Ingenieur* (Hague), vol. 64, Feb. 1, 1952, pp. E9–E17; Feb. 22, pp. E-19–E23.

519. On Magnetic Amplifiers. *Rev. Gen. de l'Électricité* (Paris), vol. 61, no. 9, 1952, p. 423.

1953

520. Transductors, **G. Barth.** *Dtsch. Electrotechnik* (Berlin), vol. 7, no. 8, Aug. 1953, pp. 387–91.

521. Dimensionless Parameters of Saturable-Reactor Magnetic Amplifiers, **V. L. Benin.** *Collected Works*, Institute of Electrical Engineering of the Acad. Sci. Ukrainian SSR, no. 10, 1953, pp. 76–87.

522. Magnetic and Dielectric Amplifiers, **G. Braun.** *VDI Zs.*, vol. 95, no. 11–12, 1953, pp. 335–40.

523. Magnetic Amplifiers, **M. Gabler.** *Elektrotech. Obs.* (Prague), an. 1, no. 1, 1953, pp. 31–39; an. 2, no. 2, pp. 103–11.

524. The Magnetic Amplifier, **W. Kafke, Y. Cenek.** *Siemens-Zs.*, vol. 27, Iss. 2, 1953, pp. 62–73.

525. Magnetic Amplifiers, **V. Kubec.** *Elektrotech. Obs.* (Prague), no. 7–8, 1953, pp. 398–415.

526. Theory of Magnetic Amplifiers with Free Magnetization, **L. Kuhn.** *Frequenz* (Berlin), vol. 7, no. 4, Apr. 1953, pp. 89–94.

527. Curves for Simultaneous Magnetization by Means of DC and AC Fields, **I. B. Negnevitsky, I. K. Panina, V. P. Mishchenko.** *Electricity* (USSR), 1953, pp. 63–64.

528. Experimental Investigation of Transient Response in Contactless Magnetic Relays, **I. B. Negnevitsky, A. L. Samurina.** *Ibid.*, no. 9, 1953.

529. Transient Response in Electrical Circuits Containing Magnetic Feedback Amplifiers, **I. D. Pashentsev.** *Collection of Scientific Work*, Leningrad Power Inst. of Railroad Eng., no. 5, 1953, pp. 74–80.

530. Magnetic Amplifiers, **M. A. Rozenblat.** *Radio Engineering* (USSR), vol. 8, no. 2, 1953, pp. 51–69.

531. Graphical Determination of Lag in Magnetic Amplifiers, **M. A. Rozenblat.** *Electricity* (USSR), no. 10, 1953, pp. 38–40.

532. Self-Oscillation in Magnetic Amplifiers, **L. V. Safris.** *J. Tech. Phys.* (USSR), vol. 23, no. 2, 1953, pp. 374–79.

533. Amplification Factor and Time Constant of the Transductor, **W. Schilling.** *EM* (Vienna), vol. 70, no. 12, 1953, pp. 261–69.

534. The Magnetic Amplifier, **W. Schilling.** *ATM* (Munich), no. 207, Apr. 1953, pp. 91–94.

535. The Dynamics of the Operation of Magnetic Amplifiers, **W. Schmidt.** *Arch. Elektr. Ubertr.* (Stuttgart), vol. 7, Iss. 5, May 1953, pp. 249–55.

536. Magnetic Amplifiers with a High Control Circuit Impedance Referred to the AC Windings, **W. Schmidt.** *Ibid.*, no. 12, pp. 574–78.

537. Design and Construction of Magnetic Amplifiers, **W. Schmidt.** *Funk und Ton* (Berlin), vol. 7, no. 7, July 1953, pp. 347–58.

538. Amplifiers without Tubes; Use of Magnetic Amplifiers for Low Frequencies, **A. M. Vincent.** *TSF et TV* (France), vol. 29, no. 291, Jan. 1953, pp. 5–7.

1954

539. Transductors, **J. Cenek.** *Elektrotechnik* (Prague), V. R. 9, C. 5, 1954, pp. 145–54.

540. Magnetic Amplifiers, **P. Colombani.** *Électricité* (Paris), vol. 38, pt. III, no. 203, 1954, pp. 61–69.

541. The Basic Circuits of the Magnetic Amplifier, **W. Dhen.** *ETZ* (Berlin), vol. 75, no. 17, Ausg. A, Sept. 1954, pp. 541–47.

542. The Dynamic Characteristics of Magnetic Amplifiers, **H. Dietz.** *VDE-Fachberichte*, S. III/28-III/45, 1954.

543. Magnetic Amplifiers, **S. Duinker.** *Tijdschr.*, Amsterdam, Holland, vol. 19, Mar. 1954, pp. 91–114.

544. Magnetic Amplifier with a Broad Pass Band, **L. Gronlie.** *Elektrotek. T.* (Oslo), vol. 67, May 25, 1954, pp. 205–08.

545. Determination of the Optimum Core Dimensions and the Number of Turns for the Saturable Reactor Control Windings in an Induction-Motor Electric Drive, **M. V. Gubanov.** *Automation and Remote Control* (USSR), vol. 15, no. 2, 1954, pp. 155–66.

546. Magnetic Amplifiers for Weak Currents, **R. Klein.** *Bull. Soc. Franc. Électr.*, vol. 4, no. 47, ser. 7, 1954, pp. 649–74.

547. Magnetic Amplifiers with Internal Feedback, **O. Kutal.** *Slaboproudy Obzor* (Prague), vol. 15, no. 11, 1954, pp. 519–27.

548. Magnetic Amplifiers (Transductors). *Mesures* (Paris), an. 19, no. 203, 1954, p. 267.

549. Principles of Duality in the Theory of Magnetic Amplifiers, **V. F. Remnev.** *Electricity* (USSR) no. 11, 1954, pp. 76–80.

550. Magnetic Amplifiers, **M. A. Rozenblat, A. F. Senchenkov.** *Ibid.*, no. 8, pp. 67–72.

551. Magnetic Amplifiers, **M. A. Rozenblat.** *Radio Engineering* (USSR), no. 6, 1954, pp. 24–28.

552. Current Controlled Magnetic Amplifiers, **W. Schilling.** *Regelungstechnik* (Munich), Iss. 6, 1954, pp. 143–49.

553. Approximate Analytical Expressions for the Characteristics of Saturable Reactor Magnetic Amplifiers without Feedback, **N. M. Tishchenko.** *Automation and Remote Control* (USSR), vol. 15, no. 2, 1954, pp. 141–54.

554. Operation of a Magnetic Amplifier with an R-L Load Connected Through a Bridge Rectifier, **N. P. Vasyleva.** *Ibid.*, pp. 544–53.

555. Graphical Analytical Method for Investigating Magnetic Amplifiers Operating in a Stabilizer Regime, **N. P. Vasyleva.** *Ibid.*, no. 1, pp. 22–34.

1955

556. Fast Response Magnetic Amplifiers Which are Controlled by AC and DC Signals, **V. G. Antoshin.** *Trans. Taganrog Radio-Eng. Inst.*, Taganrog (USSR), 1955.

557. Magnetic Amplifiers with Internal Feedback, **J. Fournier.** *Électricité* (Paris), 1955, no. 216, pp. 123–31.

558. Magnetic Amplifiers and Their Operation, **M. Gabler.** *Dtsch. Electrotechnik* (Berlin) Vol. 9, 1955, no. 1, pp. 11–16; no. 2, pp. 48–51.

559. Self Saturating Cascade Connected, Reactors, **G. Giannelli.** *Energia Nucleare*, Milan, Italy, vol. 2, 1955, p. 239.

560. Magnetic Amplifier Terminology in German Technical Literature, **F. Iser.** *Dtsch. Electrotechnik* (Berlin), no. 6, 1955, p. 225.

561. An AC Source Which is Controlled by a Magnetic Amplifier, **H. Junior.** *AEG Mitt.* (Berlin), vol. 45, nos. 1–2, 1955, pp. 79–81.

562. Theory of New Magnetic Amplifiers With a Transient Response Time Less Than One Cycle, **J. Krusek, B. Dudas.** *Slaboproudy Obzor* (Prague), vol. 16, no. 6, 1955, pp. 302–10.

563. Problems of Magnetic Preamplifiers, **F. Kummel.** *ETZ* (Berlin), vol. 76, edition A, no. 3, 1955, pp. 113–130.

564. Experimental Analysis of Magnetic Amplifiers, **M. LaFargue, R. L. Du.** *Électronique* (France), no. 100, 1955, pp. 47–52.

565. Pulse Remagnetization of Cores with Rectangular Hysteresis Loops, **M. A. Rozenblat.** *Electricity* (USSR), no. 1, 1955, pp. 24–28.

566. Magnetic Amplifiers; Part II: A Voltage-Controlled Magnetic Amplifier, **W. Schilling.** *Regelungstechnik* (Munich), vol. 3, no. 2, 1955, pp. 28–36.

567. Magnetic Elements with Rectangular Hysteresis Loops, **V. N. Tutevich.** *Electricity* (USSR), no. 11, 1955, pp. 63–67.

568. Design of Saturable-Reactor Magnetic Amplifiers with a Complex Load, **N. P. Vasilyeva, O. A. Sedykh.** *Automation and Remote Control* (USSR), vol. 16, no. 1, 1955, 16, pp. 48–63.

569. Analysis of the Operation of Intermediate Saturating Transformers for Differential Protective Devices, **A. A. Voskresensky.** *Electrical Power Stations* (USSR), 1955, pp. 35–41.

570. An AC Magnetic Amplifier, **A. I. Zhivotkovsky, A. V. Kalugin, V. K. Labutin, V. M. Bokharev.** *Trans. Scientific Research Inst.* (USSR), no. 4, 1955, pp. 17–36.

571. Design of the Characteristics of a Magnetic Amplifier Upon Magnetization with Direct and Alternating Fields, **A. V. Basharin.** *Electricity* (USSR), No. 2, 1956.

572. Constructional Constant of Magnetic Amplifiers, **S. Y. Dunaevskii.** *Electricity* (USSR), no. 2, 1956.

573. Comparison of Certain Typical Circuits of Magnetic Amplifiers with an Inner Reverse Coupling, **A. B. Gorodetskii.** *Automation and Remote Control* (USSR), vol. 17, no. 2, 1956.

574. Graphical-Analytical Method of Design of the Characteristics of Magnetic Circuits under the Joint Influence of Direct and Alternating Fields (A Continuation), **I. Y. Lekhtman.** *Ibid.*, no. 3, 1956.

575. On the Theory of an "I-deal" Self-Saturating Magnetic Amplifier, **I. V. Negnevitskii, D. A. Lipman.** *Electricity* (USSR), no. 1, 1956.

576. Principles for the Construction of Magnetic Amplifiers with a Low Threshold of Sensitivity, **M. A. Rozenblat.** *Automation and Remote Control* (USSR), vol. 17, no. 1, 1956.

577. Design of Magnetic Amplifiers on Toroidal Cores, **O. A. Sedykh.** *Ibid.*, no. 5, 1956.

578. A Magnetic Amplifier with a Parallel Load, **G. V. Subbotina.** *Ibid.*, no. 6, 1956.

579. Characteristics of Magnetic Amplifiers with Reverse Coupling, **N. M. Tishchenko.** *Ibid.*

580. Design of Push-Pull Magnetic Power Amplifiers, **N. P. Vasilyeva.** *Ibid.*, Jan. 1956, pp. 53–65.

B. Magnetic Regulation and Servo Applications

1951

581. Method for Designing a Saturable Reactor Control Circuit for Stabilizing Rectifier Output Voltages, **M. N. Gubanov.** *Electricity* (USSR), no. 3, 1951, pp. 48–53.

1952

582. A Dual-System Electro-magnetic Voltage Corrector, **V. L. Inosov.** *Collected Works*, Institute of Electr. Eng. of the Ukrainian SSR, no. 9, 1952, pp. 139–47.

583. Monitoring a Non-Symmetrical Generator Load, **I. N. Popov, M. I. Pronnikova.** *Electrical Power Stations* (USSR), no. 6, 1952, pp. 46–47.

584. Application of Saturable Reactors Biased by a DC Magnetic Field as Current Stabilizers, **M. A. Rozenblat, N. P. Vasyleva.** *Automation and Remote Control* (USSR), vol. 13, no. 6, 1952, pp. 723–30.

585. Magnetic Amplifiers: Properties of the Basic Types Used in Control Engineering, **W. Schilling.** *ATM* (Munich), no. 197, 1952, pp. 139–42.

586. Application of Magnetic Amplifiers in Control Engineering, **W. Wagner.** *EM* (Vienna), no. 11, 1952, pp. 255–61.

587. Transductors and Magnetic Amplifiers and Their Utilization in High Current Controllers, **L. Witsiosch.** *Electrotechnika* (Budapest), vol. 46, pt. I, no. 5, 1952, pp. 155–57; pt. II, no. 6, pp. 178–86.

1953

588. Transductor Regulation of Theatre and Airport Lighting, **Angeid.** *Elektroteknikeren* (Copenhagen), June 1953.

589. The Use of the Amplidyne and the Amplistat in Industry, **M. Castel.** *Bull. Sci. Ass. Ing. Élect. Montefiore* (Liege), vol. 66, iss. 2, 1953, p. 75.

590. Application of Magnetic Amplifiers in Voltage Regulators, **G. C. Manera, F. Zaccaynini.** *Elettrotecnica* (Milan), vol. 40, Oct. 12–25, 1953, pp. 553–60.

591. Remote Start-Stop Control of an Electrical Tractor Motor, **R. M. Slavin.** *Mechanisation and Electrification of Agriculture* (USSR), no. 7, 1953, p. 73.

592. Voltage Stabilizer for AC Voltages, **A. Riedel.** *Nachrichtentechnik* (Berlin), vol. 3, Oct. 1953, pp. 460–64.

1954

593. Control of Ward Leonard Sets by Magnetic Amplifiers, **J. Alle.** *Bull. Sci. Ass. Ing. Elect. Montefiore* (Liege), Dec. 1954.

594. Use of Magnetic Amplifiers in Electric Drives with Amplidyne Control, **M. V. Belyaev.** *Electricity* (USSR), no. 3, 1954.

595. Magnetic Amplifiers and Their Applications, **F. Godard.** *Bull. Soc. Belge. Électr.*, vol. 74, no. 4, 1954, pp. 233–41.

596. Magnetic Power Amplifiers for Control and Regulation, **A. Lange.** *Regelungstechnik* (Munich), vol. 2, no. 2, pp. 45–46; also *AEG Mitt.* (Berlin), no. 3–4, 1954, pp. 112–13.

597. Magnetic Amplifiers for Controlling High Speed Passenger Elevators, **C. Sichling, H. Watzing.** *VDI Zs.* (Berlin), vol. 96, no. 11–12, Apr. 15, 1954, pp. 337–40.

598. The Magnetic Amplifier for Rapid Regulation, **C. Sichling.** *VDE-Fachberichte*, III/55–III/59, 1954.

1955

599. Regulation of Ward-Leonard Sets by Magnetic Amplifiers, **J. Alle, G. Fortpied.** *ACEC Rev.* (Belgium), no. 1, 1955, pp. 27–35.

600. Decision of the Conference on Contactless Magnetic Elements and Their Applications. *Automation and Remote Control* (USSR), vol. 16, 1955, pp. 406–08.

601. A Voltage Regulator with a Magnetic Amplifier and Added Compensation for Use with Single Phase and Three Phase Generators Having Rotary Exciters, **W. Blase.** *AEG Mitt.* (Berlin), vol. 45, no. 1–2, 1955, pp. 63–69.

602. Magnetic Power Amplifiers of the "Mutrou" Type and Their Application in Control Units, Particularly for the Purposes of Voltage and Frequency Regulation of Small Generators, **M. M. Butel.** *Bull. Soc. Franc. Électr.* (Paris), vol. 5, no. 58, 1955, pp. 689–99.

603. Magnetic Amplifier for Voltage Regulation of Selenium Rectifiers, **F. Germann, E. Schröter.** *AEG Mitt* (Berlin), vol. 45, no. 1–2, 1955, pp. 85–90.

604. Static Power Converters, **S. A. Ginzburg.** *Automation and Remote Control* (USSR), vol. 16, no. 2, 1955, pp. 172–83.

605. Controlling Rectifiers by Means of Magnetic Electronic Amplifiers, **E. Golde, W. Jentsch.** *AEG Mitt.* (Berlin), vol. 45, no. 1–2, 1955, pp. 91–96.

606. Alternating Current Supply Controlled by Magnetic Amplifier, **H. Junior.** *Ibid.*

607. The Saturable Reactor and Its Application, **V. Kübec, J. Cenek.** *Dtsch. Electrotechnik* (Berlin), vol. 9, III, 3, 1955, pp. 87–92.

608. Improvement of the Operating Characteristics of a Capacitor-Excited Induction Generator by Saturated Reactors, **F. Kummel.** *Siemens-Zs.*, vol. 29, no. 8, July 1955, pp. 333–41.

609. The Possibilities of Utilizing Magnetic Amplifiers, **A. Lange.** *AEG Mitt.* (Berlin), vol. 45, no. 1–2, 1955, pp. 197–203.

II: Foreign Language Articles

B. Magnetic Regulation and Servo Applications

1955 (cont'd)

610. The Development of Voltage Regulators Which Use Magnetic Amplifiers, P. Marsili. *Elettrotecnica* (Milan), vol. 42, no. 3, 1955, pp. 165–67.

611. Controllers and Actuating Units Which are Controlled by Magnetic Amplifiers. *Mesures* (Paris), vol. 20, no. 214, 1955, pp. 197–200.

612. A Magnetic-Amplifier Controller System for Charging Large Batteries, H. F. Steinmüller. *AEG Mitt.* (Berlin), vol. 45, no. 1–2, 1955, pp. 96–99.

613. Conference on Contactless Magnetic Elements and Their Application, N. P. Vasilyeva, M. A. Boyarchenkov. *Automation and Remote Control* (USSR), vol. 16, no. 4, 1955, pp. 403–06.

614. A Comparison Between a Tirrill Voltage Regulator and a Magnetic Amplifier Type Voltage Regulator, W. Walter. *ETZ* (Berlin), vol. 76, edition A, no. 9, May 1955, pp. 327–30.

615. Stability of Generator-Voltage Control by Means of Magnetic Amplifiers When the Load is Removed, E. A. Yakubaitis. *Bull. Acad. Sci. Latvian SSR*, no. 9, 1955, pp. 87–95.

616. Static Characteristic of the Metering Unit in a Saturable-Reactor Controller, E. A. Yakubaitis. *Problems in Power Electricity*, III, Acad. Sci. of the Latvian SSR, Riga, 1955, pp. 43–61.

1956

617. An Electrical Control Device with a Regulated Speed, O. I. Aven, E. D. Demidenko, S. M. Domanitskii, E. K. Krug. *Automation and Remote Control* (USSR), vol. 17, no. 2, 1956.

618. A Two-Cycle Magnetic Amplifier with DC Out-put, R. Kh. Bal'yan. *Ibid.*

619. A Regulated Asynchronous Electric Drive with Self-Saturating Magnetic Amplifiers, M. G. Chilikin, M. M. Sokolov, V. I. Klyuchev. *Electricity* (USSR), no. 1, 1956.

620. Quick-Acting Magnetic Amplifier for Regulating Systems with Alternating Current Motors, A. I. Dem'yanckik. *Automation and Remote Control* (USSR), vol. 17, no. 3, 1956.

C. Magnetic Frequency Conversion

1951

621. On the Theory of Multiphase Frequency Multipliers, L. L. Rozhanskii. *Electricity* (USSR), no. 5, 1951.

622. A Low Frequency Oscillator, K. G. Mityushkin. *Automation and Remote Control* (USSR), vol. 12, no. 6, 1951, pp. 495–505.

1954

623. Frequency Division in the Simplest Circuit Containing Steel, S. S. Shur. *Electricity* (USSR), no. 11, 1954, pp. 52–57.

D. Operational and Instrument Magnetic Amplifiers

1952

624. Design and Calculation of Magnetic Amplifiers for the Measurement of Voltages and Low-Power Currents, I. K. Parra. *Collected Works*, Electr. Eng. Inst. of the Acad. Sci. Ukrainian SSR, no. 9, 1952, pp. 57–70.

625. Direct Current Measuring Transformers, Yu. G. Tol'stov. *Gosenergoisdat*, Moscow, USSR, 1951.

1953

626. The Problems of Designing Ferromagnetic Metering Amplifiers, E. K. Mikhailova. *Transactions*, All-Union Scientific Research Inst. of Metrology, no. 14, 1953, p. 100.

1954

627. Magnetometers: 1. Definition of Concepts and Theory. 2. Devices for Measuring the Earth's Magnetic Field. *ATM* (Munich), no. 220, 1954, pp. 115–18.

628. Amplification of Small AC Voltages by Means of Magnetic Amplifiers, F. Kummel. *ETZ* (Berlin), Ausg. A, vol. 75, June 1954, pp. 367–79.

629. Design of DC Metering Transformers, A. M. Ryvkin. *Electricity* (USSR), no. 9, 1954, pp. 26–28.

630. Magnetic Amplifier as a Phase Detecting Device, Y. Sakurai, H. Nagano. *Tech. Repts.*, Osaka Univ., Osaka, Japan, vol. 4, no. 114, Oct. 1954, pp. 221–26.

1955

631. Transductors for Measurement, L. Borg. *ASEA J.*, vol. 28, no. 7, 1955, pp. 110–16.

632. Transductor for Measuring Large Direct Currents, E. Tomanek, B. B. Dudas, J. Krusek. *Elektrotechnik* (Prague), vol. 10, no. 3, 1955, p. 74–77.

633. Magnetic Amplifier Type Recording Instrument of Electrolysis Study, J. Yamaguchi, Y. Sakurai. *Tech. Repts.*, Osaka Univ., Osaka, Japan, vol. 5, 1955.

1956

634. Use of Magnetic Zero-Instruments in Telemetering and Recording Devices, A. M. Pshenichnikov. *Electricity* (USSR), no. 1, 1956.

E. Static Magnetic Switching Applications

1951

635. Sensitive Current Relays for Protection from Shorts to Ground, and the Technique of Their Application, Y. S. Gelfand, I. N. Popov, G. V. Subbotina. *Electrical Power Stations* (USSR), no. 11, 1951, pp. 41–43.

1953

636. Saturating Current Transformer With Short-Circuited Turns Utilized in Protected Relay Circuits, D. A. Drozdov. *Electricity* (USSR), no. 10, 1953, pp. 32–36.

637. Push-Pull Magnetic Trigger, P. V. Karpenko. *Trans. Moscow Power Inst.*, no. 14, 1953, pp. 63–72.

638. Contactless Magnetic Time Relay, I. S. Mikadze, I. B. Negnevitsky. *Ibid.*, pp. 53–62.

639. Application of Transformer Relays for the Differential Protection of Power Transformers, V. V. Mikhailov. *Trans. Novocherkassk Poly. Inst.*, Novocherkassk, USSR, vol. 23, 1953.

640. Experimental Investigation of Transient Response in Contactless Magnetic Relays, I. B. Negnevitsky, A. L. Samurina. *Electricity* (USSR), no. 9, 1953, pp. 64–69.

1954

641. A Contactless Magnetic Relay, M. A. Rozenblat, S. A. Dokhman. *Automation and Remote Control*. (USSR), vol. 15, no. 1, 1954, pp. 3–21.

642. A Commutator Made on Magnetic Elements with Rectangular Hysteresis Loops, V. N. Tutevich, V. A. Zhozhikashvili. *Automation and Remote Control* (USSR), vol. 15, no. 1, 1954.

1955

643. The Operation of Computing Switching Circuits with Magnetic Elements Having Hysteresis Loops, V. A. Zhozhikashvili, G. Mityushkin. *Automation and Remote Control* (USSR), vol. 16, no. 4, 1955.

1952

644. Application of Modern Ferro-Magnetic Alloys, J. Moline. *Radio Francaise* (Paris), no. 5, May 1952, pp. 10–13.

1953

645. Certain Properties of a Permalloy of Thin Cross Section, L. I. Rabkin, P. I. Yuzvinskaya. *Electricity* (USSR), no. 10, 1953.

1955

646. Ribbon Cores for Transformers, F. W. Neumann. *Nachrichtentechnik* (Berlin), vol. 5, no. 5, 1955, pp. 225–28.

647. New Magnetic Materials with Oriented Crystals, G. H. Perrin. *Bull. Soc. Franc. Électr.* (Paris), vol. 5, Ser. 7, no. 53, 1955, pp. 305–18.

1956

648. Magnetic Alloys for Cores of Magnetic Amplifiers and Their Curves of Simultaneous Magnetization with Direct and Alternating Fields, O. I. Aven. *Automation and Remote Control* (USSR), vol. 17, no. 4, 1956.

H. Measurements and Testing

1951

649. The Vector Meter; Measurement of Loss Factor in Whole Sheets of Transformer Steel, F. Koppelmann. *EM* (Vienna), May 1951, pp. 225–29.

1955

650. Testing Ferromagnetic Materials, I. I. Kifer V. S. Pantyushin. *Magnetic Measurements* (State Power Press, USSR), 1955, p. 240.

651. Plotting Magnetization Curves and Hysteresis Loops, H. Neumann. *ATM* (Munich), May 1955, pp. 105–08.

652. Impulse Magnetization of Cores with Rectangular Hysteresis Loops, M. A. Rozenblat. *Electricity* (USSR), no. 1, 1955.

653. Measurement of Properties of Magnetically Soft Materials by the Bridge Method, E. I. Gurevich. *Ibid.*

J. Associated Devices

1952

654. Operation of Semi-Conductor Rectifiers in Control Circuits, V. G. Komar. *Gosenergoisdat*, Moscow, USSR, 1952.

1956

655. Stabilized Selenium Rectifier Devices, I. I. Patgauz. *Electricity* (USSR), no. 4, 1956.

K. Ferroresonance and Its Applications

1951

656. Calculation of Quasi-Relaxational Oscillations in Ferro-Resonant Circuits Containing Saturable Reactors, V. E. Bogolyubov. *Electricity* (USSR), no. 8, 1951, pp. 64–68.

L. Manufacturing Techniques, Environmental Effects, and Reliability

1956

657. Interdepartmental Conference on the Technology of Manufacturing Magnetic Amplifiers. *Automation and Remote Control* (USSR), vol. 17, no. 10, Oct. 1956, pp. 943–45.

III: Domestic Patents

A. Theory, Analysis, and Design

1951

658. Voltage Doubler Magnetic Amplifier, R. E. Morgan. *No. 2,552,203*, Apr. 2, 1948–May 8, 1951.

659. **Magnetic Amplifier, P. Gachet, P. J. Augier de Montgremier.** *No. 2,552,952*, July 15, 1948–May 15, 1951.

660. **Magnetic Amplifier Control System, R. E. Morgan.** *No. 2,554,203*, Nov. 10, 1949–May 22, 1951.

661. **Magnetic Amplifier, G. Hepp.** *No. 2,556,083*, July 8, 1946–June 5, 1951.

662. **Means for Reinforcing Electric Currents by Direct-Current Saturated Inductances, U. Lamm.** *No. 2,562,047*, Oct. 21, 1947–July 24, 1951.

663. **Three Phase Control System Employing Saturable Reactors, R. E. Morgan.** *No. 2,567,725*, Oct. 31, 1950–Sept. 11, 1951.

664. **Means for Reproducing a Direct Current, H. Forssel.** *No. 2,573,255*, Dec. 24, 1947–Oct. 30, 1951.

665. **Magnetic Amplifier, E. G. Downie.** *No. 2,578,405*, Feb. 28, 1950–Dec. 11, 1951.

1952

666. **Electromagnetic System, A. S. FitzGerald.** *No. 2,584,856*, Dec. 31, 1946–Feb. 5, 1952.

667. **Spare Connected Polyphase Transductor, H. M. Sörensen.** *No. 2,599,137*, Apr. 20, 1949–June 3, 1952.

668. **Transductor, J. W. Trindle.** *No. 2,603,768*, Apr. 20, 1950–July 15, 1952.

669. **Saturable Reactor Apparatus, H. M. Ogle.** *No. 2,617,090*, June 17, 1950–Nov. 4, 1952.

670. **Direct Current Control System, M. Bracutt.** *No. 2,622,239*, Mar. 18, 1950–Dec. 16, 1952.

1953

671. **Magnetic Amplifier Circuits for Applying Reversible Direct-Current Voltage to Inductive Loads, W. F. Eames, W. H. Esselman.** *No. 2,629,847*, Aug. 25, 1950–Feb. 24, 1953.

672. **Magnetic Amplifier System, H. F. McKenney, H. Kaplan.** *No. 2,636,150*, Mar. 30, 1951–Apr. 21, 1953.

673. **Spare Connected Polyphase Transductor, H. M. Sörensen, H. Forssel.** *No. 2,640,963*, June 14, 1949–June 2, 1953.

674. **Separate Magnetization of Magnetic Amplifiers, R. A. Ramey.** *No. 2,644,129*, Dec. 27, 1950–June 30, 1953.

675. **Electromagnetic Audio Amplifier, W. C. Kluz.** *No. 2,657,281*, Feb. 15, 1950–Oct. 27, 1953.

676. **Saturable Core Transformer System, A. V. Hemingway, P. D. Atkinson.** *No. 2,661,453*, Aug. 3, 1949–Dec. 1, 1953.

1954

677. **Magnetic Amplifier Circuit, R. E. Morgan.** *No. 2,675,518*, Feb. 28, 1951–Apr. 13, 1954.

678. **Reversible Inductive Load Controlled Magnetic Amplifiers, J. T. Carleton.** *No. 2,677,084*, Apr. 9, 1952–Apr. 27, 1954.

679. **Magnetic Amplifier, F. J. Rau.** *No. 2,677,099*, Mar. 29, 1952–Apr. 27, 1954.

680. **Saturable Transformer Device, A. Bennett.** *No. 2,678,419*, Sept. 30, 1946–May 11, 1954.

681. **Magnetic Amplifier Circuit, R. Cohen, H. F. Storm.** *No. 2,682,632*, May 20, 1949–June 29, 1954.

682. **Power Transmission, F. G. Logan.** *No. 2,683,853*, Nov. 29, 1951–July 13, 1954.

683. **Magnetic Amplifier System, D. T. Kadushin, H. Kaplan.** *No. 2,688,723*, Sept. 16, 1949–Sept. 7, 1954.

684. **Magnetic Amplifier, W. H. Newell.** *No. 2,688,724*, Apr. 27, 1951–Sept. 7, 1954.

685. **Saturable Reactor Device, F. G. Logan.** *No. 2,689,328*, Dec. 1, 1950–Sept. 14, 1954.

686. **Magnetic Amplifier System, J. Stone.** *No. 2,697,813*, Apr. 12, 1951–Dec. 21, 1954.

1955

687. **Magnetic Amplifier System, H. F. Storm.** *No. 2,704.823*, Feb. 29, 1952–Mar. 22, 1955.

688. **Electromagnetic Audio Amplifiers, F. G. Logan.** *No. 2,710,313*, Oct. 12, 1948–June 7, 1955.

689. **Electric Motor Control Circuit Using Saturable Reactors, R. E. Morgan.** *No. 2,715,203*, June 17, 1950–Aug. 9, 1955.

690. **Magnetic Amplifier with High Gain and Rapid Response, R. A. Ramey.** *No. 2,719,885*, July 20, 1951–Oct. 4, 1955.

691. **Magnetic Amplifier for Power Transmission, S. Steinitz.** *No. 2,723,373*, Apr. 25, 1950–Nov. 8, 1955.

1956

692. **Magnetic Amplifier, F. H. Beisey.** *No. 2,730,574*, Aug. 22, 1951–Jan. 10, 1956.

693. **Magnetic Amplifiers, J. A. Crawford.** *No. 2,731,521*, Aug. 12, 1952–Jan. 17, 1956.

694. **Rapid Response Self-Saturating Magnetic Amplifier, B. D. Bedford.** *No. 2,733,306*, July 29, 1952–Jan. 31, 1956.

695. **Stabilized Full-Wave Self-Saturating Magnetic Amplifier, H. M. Ogle.** *No. 2,733,307*, Apr. 23, 1953–Jan. 31, 1956.

696. **Magnetic Amplifier, A. W. Rechten, B. M. Bellman.** *No. 2,733,388*, Feb. 18, 1952–Jan. 31, 1956.

697. **Self-Excited Transductors, A. Arvidsson.** *No. 2,733,403*, June 16, 1952–Jan. 31, 1956.

698. **Magnetic Amplifier with Half-Wave Phase-Reversal Type Output, C. W. Lufcy, A. E. Schmid, Jr.** *No. 2,734,165*, June 30, 1952–Feb. 7, 1956.

699. **Magnetic Amplifier Circuit, M. B. Zucchino.** *No. 2,745,056*, May 4, 1953–May 8, 1956.

700. **Magnetic Amplifier, S. B. Cohen, J. Bentkowsky.** *No. 2,745,908*, Jan. 10, 1951–May 15, 1956.

701. **Magnetic Amplifier, K. M. Hanson.** *No. 2,752,429*, Jan. 7, 1952–June 26, 1956.

702. **Magnetic Amplifiers, F. A. Baker.** *No. 2,752,560*, Apr. 18, 1955–June 26, 1956.

703. **Magnetic Amplifier, C. J. Creveling.** *No. 2,753,518*, June 1, 1953–July 3, 1956.

704. **Half-Wave Bridge Magnetic Amplifier, E. T. Hooper, Jr.** *No. 2,754,473*, Apr. 9, 1953–July 10, 1956.

705. **Arrangement for Producing Full-Wave Output from Half-Wave Magnetic Amplifiers, P. W. Barnhart.** *No. 2,754,474*, Apr. 13, 1955–July 10, 1956.

B. Magnetic Regulation and Servo Applications

1951

706. **Follow-Up Apparatus, G. N. Burkhart, Jr.** *No. 2,537,027*, Feb. 8, 1947–Jan. 9, 1951.

707. **Voltage Regulator, M. Ker.** *No. 2,541,820*, Nov. 12, 1946–Feb. 13, 1951.

708. **Saturable Core Reactor, B. D. Bedford.** *No. 2,547,615*, Oct. 26, 1948–Apr. 3, 1951.

709. **Magnetic Amplifier, R. W. Olson.** *No. 2,548,049*, July 10, 1948–Apr. 10, 1951.

710. **Heat-Regulating System, A. S. FitzGerald.** *No. 2,554,945*, July 11, 1947–May 29, 1951.

711. **Motor Control System, H. M. Ogle.** *No. 2,555,992*, Feb. 3, 1950–June 5, 1951.

712. **Electric Controlling Apparatus, F. G. Logan.** *No. 2,558,572*, June 9, 1945–June 26, 1951.

713. **Servo System with Saturable Core Component, J. P. Palmer.*** *No. 2,559,513*, Dec. 28, 1948–July 3, 1951.

714. **Reversing Magnetic Amplifier Control System, W. L. O. Graves.** *No. 2,571,708*, May 21, 1949–Oct. 16, 1951.

715. **Saturable Reactor Impedance Relay, F. Dahlgren.** *No. 2,573,249*, Aug. 23, 1947–Oct. 30, 1951.

1952

716. **Electric Controlling Apparatus, F. G. Logan.** *No. 2,585,332*, Feb. 10, 1948–Feb. 12, 1952.

717. **Rectifier Connection, H. Öfverholm, H. Forssell.** *No. 2,588,155*, Sept. 10, 1947–Mar. 4, 1952.

718. **Magnetic Amplifier Relay Circuits, W. F. Horton.** *No. 2,594,022*, Aug. 5, 1950–Apr. 22, 1952.

719. **Regulator with Saturable Core, S. E. Hedström.** *No. 2,596,685*, Sept. 17, 1947–May 13, 1952.

720. **Voltage Regulator, H. M. Ogle.** *No. 2,615,155*, Oct. 21, 1952.

1953

721. **Saturable Core Reactor System, F. Landau, J. E. Carlsson, S. E. Hedström.** *No. 2,625,677*, Aug. 21, 1947–Jan. 13, 1953.

722. **Regulating System, E. L. Harder, S. L. Bradley.** *No. 2,629,853*, Dec. 22, 1951–Feb. 24, 1953.

723. **Voltage Regulator, D. H. Ransom, H. M. Miller.** *No. 2,631,268*, Mar. 13, 1951–Mar. 10, 1953.

724. **Self-Biasing Magnetic Amplifier, D. J. Sikorra.** *No. 2,632,145*, Jan. 20, 1951–Mar. 17, 1953.

725. **Voltage Regulating Device, Y. A. Rocard.** *No. 2,633,562*, Oct. 29, 1949–Mar. 31, 1953.

726. **Electrical Control Circuit, P. Glass.** *No. 2,634,395*, July 12, 1945–Apr. 7, 1953.

727. **Regulating System, S. L. Bradley.** *No. 2,644,127*, Dec. 22, 1951–June 30, 1953.

728. **Stabilizer for Alternating Current Power Transmission Systems, E. F. W. Alexanderson, R. W. Kuenning.** *No. 2,644,898*, Feb. 28, 1951–July 7, 1953.

729. **Voltage Regulators, F. R. Milsom.** *No. 2,651,020*, July 23, 1951–Sept. 1, 1953.

1954

730. **Magnetic Amplifier Motor Control Device, F. S. Malick.** *No. 2,675,513*, June 4, 1953–Apr. 13, 1954.

731. **Magnetic Amplifier for Controlling the Voltages Applied to Motor Armatures, F. S. Malick.** *No. 2,677,088*, Aug. 9, 1952–Apr. 27, 1954.

732. **Regulator System for Generators, J. T. Carleton.** *No. 2,677,097*, Feb. 5, 1953–Apr. 27, 1954.

733. **Two-Phase Induction Motor Magnetic Amplifier with Direct-Current Braking, W. A. Geyger.** *No. 2,677,796*, Mar. 11, 1952–May 4, 1954.

734. **Inverter Frequency Regulator, H. H. Woerdemann.** *No. 2,682,633*, Apr. 5, 1952–June 29, 1954.

735. **Magnetic Servo Amplifier with Direct-Current Damping, W. A. Geyger.** *No. 2,683,843*, July 20, 1953–July 13, 1954.

736. **Bridge Control Circuits, W. H. Ingersoll.** *No. 2,691,130*, Aug. 11, 1952–Oct. 5, 1954.

737. **Saturable Reactor Using Permanent Magnets, W. J. Brown.** *No. 2,696,583*, Aug. 17, 1949–Dec. 7, 1954.

738. **Saturable Reactor Motor Control Systems, W. J. Brown.** *No. 2,697,194*, Aug. 17, 1949–Dec. 14, 1954.

739. **Control System Using Multiple Winding Saturable Reactors, W. J. Brown.** *No. 2,697,197*, Aug. 17, 1949–Dec. 14, 1954.

1955

740. **Voltage Regulator Employing Saturable Reactor, R. E. Morgan.** *No. 2,699,525*, Feb. 8, 1951–Jan. 11, 1955.

741. **Magnetic Voltage Regulator, H. H. Woerdemann.** *No. 2,700,128*, May 5, 1952–Jan. 18, 1955.

742. **Motor Control Systems, N. S. Wells.** *No. 2,709,774*, Nov. 13, 1952–May 31, 1955.

743. **Alternating-Current Motor, E. F. W. Alexanderson.** *No. 2,711,502*, July 21, 1954–June 21, 1955.

744. **Alternating Current Motor Control,**

III: Domestic Patents

B. Magnetic Regulation and Servo Applications

1955 (cont'd.)

G. E. Mathias, W. T. Hunt. *No. 2,712,105*, Nov. 19, 1952–June 28, 1955.

745. Electric Regulation by Controlled Reversible Motor, **W. Few.** *No. 2,713,137*, Oct. 31, 1949–July 12, 1955.

746. Regulating System with Exciter Field Winding in Series with Alternator Field Winding, **W. L. Ringland.** *No. 2,715,205*, May 18, 1953–Aug. 9, 1955.

747. Magnetic Servo Amplifier, **W. A. Geyger.** *No. 2,717,346*, Apr. 2, 1953–Sept. 6, 1955.

748. Amplifier Limit Circuit, **V. J. Louden.** *No. 2,717,355*, Aug. 18, 1952–Sept. 6, 1955.

749. Alternating Current Motor Control Systems for Printing Presses or the Like, **H. R. Behr, E. C. Rhyne.** *No. 2,719,255*, Oct. 6, 1951–Sept. 27, 1955.

750. Voltage-Controlled A C Generator System, **D. G. Scorgie.** *No. 2,719,260*, May 29, 1953–Sept. 27, 1955.

751. Temperature Control System for Electric Heating Elements, **R. E. Morgan.** *No. 2,720,579*, Apr. 3, 1952–Oct. 11, 1955.

752. Motor Control Apparatus, **P. Siskind.** *No. 2,721,298*, Sept. 25, 1951–Oct. 18, 1955.

753. Voltage Regulators, **N. Silver.** *No. 2,721,303*, Nov. 22, 1954–Oct. 18, 1955.

754. Voltage Regulator, **N. Silver, D. W. Tanner.** *No. 2,721,304*, June 21, 1954–Oct. 18, 1955.

755. Current Limit Motor Control System, **W. T. Hunt, Jr.** *No. 2,722,642*, Jan. 30, 1953–Nov. 1, 1955.

756. Control System for Alternating Current Motors, **E. C. Rhyne, Jr.** *No. 2,722,643*, Oct. 6, 1951–Nov. 1, 1955.

757. Alternating-Current Motor Control Systems, **E. C. Rhyne, H. R. Behr.** *No. 2,722,646*, Oct. 6, 1951–Nov. 1, 1955.

758. Regulating System Utilizing a Saturable Reactor Having Negative Feedback, **D. J. Sikorra.** *No. 2,722,654*, Dec. 11, 1950–Nov. 1, 1955.

759. Line Voltage Regulator, **F. W. Anderson.** *No. 2,722,655*, Apr. 10, 1953–Nov. 1, 1955.

760. Stabilizing Circuit for Alternating-Current Generator, **H. F. Storm.** *No. 2,724,797*, Feb. 1, 1951–Nov. 22, 1955.

761. Frequency-Selective System, **H. B. Rex.** *No. 2,724,807*, Feb. 16, 1950–Nov. 22, 1955.

762. Magnetic Amplifier Electrical Position Control System, **F. S. Malick, C. L. Mershon.** *No. 2,725,519*, Mar. 31, 1952–Nov. 29, 1955.

763. Electrical Error Detector, **W. H. Woodworth.** *No. 2,725,520*, Nov. 4, 1954–Nov. 29, 1955.

764. Differential Coupling Circuit for Multi-Stage Half-Wave Magnetic Servo Amplifiers, **W. A. Geyger.** *No. 2,725,521*, Jan. 26, 1955–Nov. 29, 1955.

765. Variable Frequency Power Control Systems for A-C Motors, **H. M. Ogle.** *No. 2,727,199*, May 18, 1953–Dec. 13, 1955.

766. Motor Control System, **M. A. Edwards.** *No. 2,727,201*, July 29, 1952–Dec. 13, 1955.

767. Regulator Systems, **G. H. Stearley.** *No. 2,728,044*, Aug. 16, 1954–Dec. 20, 1955.

1956

768. Magnetic Amplifier Circuit for Motor, **E. L. Gardner.** *No. 2,729,777*, Nov. 12, 1954–Jan. 3, 1956.

769. Motor Control System Utilizing Magnetic Amplifiers, **F. R. Milsom.** *No. 2,729,779*, Dec. 12, 1950–Jan. 3, 1956.

770. Electric Control Apparatus, **A. H. B. Walker, K. G. King.** *No. 2,732,505*, Aug. 26, 1953–Jan. 24, 1956.

771. Thermal Time-Constant Circuit, **H. M. Ogle.** *No. 2,733,404*, Nov. 27, 1951–Jan. 31, 1956.

772. Three Phase Motor System, **H. F. McKenney, L. H. Rowley.** *No. 2,734,157*, Jan. 15, 1953–Feb. 7, 1956.

773. Magnetic Amplifier Regulated Current Limit Control for Ward Leonard Systems, **C. H. Storey, Jr., W. T. Hunt, Jr.** *No. 2,735,053*, Feb. 27, 1953–Feb. 14, 1956.

774. Magnetic Amplifier Motor Speed Control, **F. S. Malick.** *No. 2,735,060*, Aug. 16, 1952–Feb. 14, 1956.

775. Control System Using Two Saturable Reactors in a Constant Current Circuit, **R. L. Peterson.** *No. 2,740,933*, Jan. 26, 1953–Apr. 3, 1956.

776. Voltage Stabilizer and Reference Circuit Therefor, **R. E. Anderson.** *No. 2,740,936*, Dec. 30, 1953–Apr. 3, 1956.

777. Voltage Reference Circuit, **J. T. Carleton.** *No. 2,743,152*, Mar. 26, 1953–Apr. 24, 1956.

778. Motor Control by Magnetic Servo Amplifier, **W. A. Geyger.** *No. 2,743,407*, Nov. 30, 1953–Apr. 24, 1956.

779. Load and Control Current Feedback Magnetic Amplifier, **H. H. Woerdemann.** *No. 2,745,055*, Dec. 19, 1951–May 8, 1956.

780. Motor Control System, **R. S. Crenshaw, Jr.** *No. 2,750,549*, June 1, 1953–June 12, 1956.

781. Magnetic Type Temperature Controller, **W. T. Hage.** *No. 2,752,473*, June 26, 1952–June 26, 1956.

782. Control System Utilizing Magnetic Amplifier for Reference Voltage, **A. C. Halter.** *No. 2,752,545*, Nov. 18, 1953–June 26, 1956.

783. Magnetic Amplifier Motor Control, **E. F. W Alexanderson.** *No. 2,752,549*, July 27, 1953–June 26, 1956.

784. Adjustable Voltage Drives for Electric Motor, **E. Pell.** *No. 2,753,504*, Nov 22, 1954–July 3, 1956.

785. Combination Control and Reference Circuit for a Magnetic Amplifier, **E. T. Hooper, Jr.** *No. 2,753,517*, Mar. 27, 1953–July 3, 1956.

C. Magnetic Frequency Conversion

1952

786. Harmonic Generator, **C. A. Lovell, D. B. Parkinson.** *No. 2,580,446*, Dec. 31, 1948–Jan. 1, 1952.

787. Frequency Generator Employing a Magnetic Double Cross Valve, **H. J. McCreary.** *No. 2,611,119*, Sept. 15, 1950–Sept. 16, 1952.

788. Magnetic Cross Valve Having Self Starting Characteristics, **H. J. McCreary.** *No. 2,611,120*, June 21, 1950–Sept. 16, 1952.

789. Device for the Polyphase Transformation of the Frequency of Three-phase Circuits, **U. Corbino.** *No. 2,616,070*, Dec. 8, 1949–Oct. 28, 1952.

1953

790. Electrical Converter, **A. W. La Marche.** *No. 2,651,018*, May 8, 1950–Sept. 1, 1953.

1954

791. Frequency Multiplier, **W. Krämer.** *No. 2,666,178*, Oct. 7, 1950–Jan. 12, 1954.

1955

792. Magnetic Cross Valve Circuits, **H. J. McCreary.** *No. 2,703,388*, May 16, 1950–Mar. 1, 1955.

D. Operational and Instrument Magnetic Amplifiers

1951

793. Computer Using Magnetic Amplifiers, **B. Rossi, H. S. Sack, R. T. Beyer, G. H. Miller.** *No. 2,574,438*, July 3, 1946–Nov. 6, 1951.

1952

794. Method and Apparatus for Measuring Intensity of Magnetic Fields, **G. Muffly.** *No. 2,598,285*, June 23, 1947–May 27, 1952.

1953

795. Potentiometer Measuring System and Apparatus, **A. Hansen, Jr.** *No. 2,645,748*, June 21, 1951–July 14, 1953.

796. Power Measuring Device, **T. R. Specht.** *No. 2,656,511*, Nov. 7, 1951–Oct. 20, 1953.

1955

797. Self-Balancing Magnetic Amplifier, **W. A. Geyger.** *No. 2,700,130*, July 2, 1952–Jan. 18, 1955.

1956

798. Electromagnetic Transformer, **R. W. Gilbert.** *No. 2,729,781*, Oct. 31, 1952–Jan. 3, 1956.

799. Amplitude Recording System Utilizing Saturable Reactors, **D. T. N. Williamson.** *No. 2,730,694*, Jan. 29, 1952–Jan. 10, 1956.

E. Static Magnetic Switching Applications

1953

800. Nonlinear Resonant Flip-Flop Circuit, **C. F. Spitzer, R. J. Reich.** *No. 2,653,254*, Apr. 23, 1952–Sept. 22, 1953.

1955

801. High Speed Magnetic Trigger Circuit, **R. A. Ramey.** *No. 2,717,965*, May 18, 1953–Sept. 13, 1955.

802. Counting Circuit, **C. L. Isborn.** *No. 2,721,947*, May 3, 1954–Oct. 25, 1955.

803. Circuit Means for Selecting the Highest or Lowest of a Plurality of Signals, **W. J. Dunnet.** *No. 2,725,549*, May 25, 1954–Nov. 29, 1955.

1956

804. Magnetic Flip-Flop, **J. Montner.** *No. 2,747,109*, Sept. 4, 1953–May 22, 1956.

K. Ferroresonance and Its Applications

1955

805. Non-Linear Resonant Trigger Circuits, **C. F. Spitzer, R. T. Gordon.** *No. 2,723,353*, May 29, 1953–Nov. 8, 1955.

806. Ferro-Resonant Shift Register, **C. L. Isborn.** *No. 2,723,354*, Sept. 28, 1954–Nov. 8, 1955.

1956

807. Means for Producing Low-Frequency Electrical Oscillations, **L. Borg.** *No. 2,730,670*, Nov. 7, 1950–Jan. 10, 1956.

808. Voltage Reference Devices, **W. G. Evans.** *No. 2,739,282*, Feb. 5, 1953–Mar. 20, 1956.

IV: Foreign Patents

A. Theory, Analysis, and Design

1951

809. Improvements in Saturable Core Reactance Devices, **F. H. Belsey, J. S. Wilson.** *British No. 659,762*, Nov. 24, 1948–Oct. 24, 1951.

1952

810. Electrical Control Circuit with Saturable Reactors, **R. E. Morgan.** *German No. 847,025*, June 17, 1951–Aug. 18, 1952.

811. Improvements in Saturable-Core Re-

actor Devices, F. H. Belsey, J. S. Wilson. *British No. 685,320*, Apr. 24, 1950–Dec. 31, 1952.

1953

812. Arrangement for Adjusting the Magnitude of the Output Current of a Magnetic Amplifier with No-Signal Conditions, G. Rudolph. *German No. 756,138*, Apr. 8, 1941–May 4, 1953.

813. Polarity-Sensitive Magnetic Amplifier, G. Barth. *German No. 758,289*, Aug. 20, 1938–Mar. 2, 1953.

814. Saturable-Reactor Arrangement, W. Lattemann. *German No. 863,089*, Feb. 25, 1941–Jan. 15, 1953.

815. Saturable-Reactor Device Acting as a Magnetic Amplifier, G. Rudolph. *German No. 899,370*, July 6, 1941–Dec. 10, 1953.

1954

816. Magnetic Amplifier, G. Rudolph. *German No. 902,633*, July 25, 1941–Jan. 25, 1954.

817. Saturable-Reactor Arrangement Particularly Suitable as a Magnetic Amplifier, G. Rudolph. *German No. 906,706*, Aug. 20, 1941–Mar. 18, 1954.

818. Magnetic Amplifier, G. Barth. *German No. 914,390*, Dec. 25, 1938–July 1, 1954.

819. Saturable Reactor Device Controlled by Pre-Magnetization, G. Rudolph. *German No. 914,864*, Aug. 22, 1940–July 12, 1954.

820. Saturable Reactor Device Controlled by Pre-Magnetization, G. Rudolph. *German No. 917,128*, Aug. 20, 1941–Aug. 26, 1954.

821. Magnetic Amplifier, G. Russ, H. Watzinger. *German No. 917,311*, June 6, 1952–Aug. 30, 1954.

822. Magnetic Amplifier, G. Barth. *German No. 917,374*, Aug. 1, 1942–Sept. 2, 1954.

823. Magnetic Amplifier, G. Rudolph. *German No. 918,448*, Feb. 13, 1943–Sept. 27, 1954.

824. Polarity-Sensitive Magnetic Amplifier with Feedback, G. Barth. *German No. 919,897*, Dec. 1, 1938–Nov. 8, 1954.

B. Magnetic Regulation and Servo Applications

1951

825. Magnetic Amplifier for Control Purposes, R. E. Morgan. *German No. 821,519*, May 21, 1950–Nov. 19, 1951.

1952

826. Rectifiers with Transductors for Automatic Control of the Voltage Supplied from These Rectifiers, S. E. Hedström, U. Lamm. *Swedish No. 138,376*, Nov. 16, 1945–Dec. 9, 1952.

827. Current-Limiting Device for a Rectifier, N. O. H. Berglund. *Swedish No. 138,576*, July 20, 1949–Dec. 30, 1952.

828. Device for Current-Limiting on a Transductor-Controlled Rectifier, N. O. H. Berglund. *Swedish No. 138,579*, Mar. 14, 1950–Dec. 30, 1952.

1953

829. Arrangement with D-C Pre-Magnetized Saturable Reactors and D-C Controlled Magnetic Amplifier and Regulator Device, P. H. E. Claesson. *Swedish No. 139,069*, Nov. 11, 1944–Feb. 10, 1953.

830. Arrangement for Magnetic Amplifiers with Inductive Load, K. Pfister. *German No. 896,666*, Feb. 17, 1951–Nov. 12, 1953.

1954

831. Regulator Device, K. G. King. *Swedish No. 145,762*, Oct. 31, 1952–June 15, 1954.

832. Regulator Device for Rectifiers, K. G. Lindholm. *Swedish No. 145,851*, Dec. 12, 1947–June 22, 1954.

833. Arrangement for Voltage-Control and Current-Limiting by Means of a Transductor-Controlled Rectifier, A. Köhler. *Swedish No. 147,683*, Jan. 29, 1951–Nov. 16, 1954.

1955

834. Arrangement for Controlling a Rectifier by Means of a Transductor, A. Arvidsson. *Swedish No. 149,569*, Nov. 17, 1948–Apr. 12, 1955.

D. Operational and Instrument Magnetic Amplifiers

1953

835. Static D-C Instrument Transformer with Polarity-Reversible Output Current, W. Krämer. *German No. 866,066*, May 3, 1950–Feb. 5, 1953.

836. Magnetic Amplifier for Low-Level Input-Power Values, H. Neidhardt. *German No. 877,917*, Sept. 5, 1951–May 28, 1953.

837. D-C Instrument Transformer for Measurement of High D-C Voltages and High-Voltage Direct Currents, S. Nowitzki, W. Hartel. *German No. 882,272*, Dec. 24, 1943–July 6, 1953.

838. Pre-Magnetized Saturable Reactor, Particularly D-C Instrument Transformer with Polarity-Reversible Arrangement, W. Blankenburg. *German No. 891,887*, Feb. 2, 1943–Oct. 1, 1953.

839. Arrangement for Reproducing a Direct Current, H. Forssel. *German No. 892,476*, June 1, 1943–Oct. 8, 1953.

840. Arrangement for Measuring Large Direct Currents, W. A. Geyger. *German No. 893,091*, Mar. 7, 1941–Oct. 12, 1953.

841. Arrangement for Measurement of Direct Currents, W. Kramer. *German No. 897,447*, Oct. 8, 1936–Oct. 15, 1953.

842. Arrangement for Measurement, Indication, and Control of Direct Currents and D-C Voltages, H. Collins. *German No. 900,967*, Dec. 18, 1941–Nov. 19, 1953.

1954

843. Instrument Transformer for Measurement of Very High D-C Voltages, W. Krämer. *German No. 905,407*, Oct. 18, 1941–Jan. 14, 1954.

844. D-C Instrument Transformer, W. Krämer. *German No. 907,673*, Oct. 14, 1939–Feb. 18, 1954.

845. D-C Instrument Transformer, W. Krämer. *German No. 901,454*, Jan. 15, 1938–Mar. 25, 1954.

846. D-C Instrument Transformer with Negative-Feedback Ampere-turns, W. Blankenburg. *German No. 913,198*, Mar. 6, 1943–June 10, 1954.

1955

847. Arrangement for Measuring Direct Currents and D-C Voltages, W. A. Geyger. *German No. 932,508*, Dec. 15, 1938–Sept. 1, 1955.

848. Arrangement for Measuring Small D-C Quantities by Means of an Alternating-Current Differential-Type Circuit, W. A. Geyger. *German No. 935,141*, Nov. 12, 1939–Nov. 10, 1955.

1956

849. Arrangement for Reducing the Response Time of the Electrical Measuring Instrument in Self-Balancing Potentiometer Devices, W. A. Geyger. *German No. 940,177*, Dec. 13, 1939–Mar. 15, 1956.

The Analysis of Redundancy Networks

FRED MOSKOWITZ
NONMEMBER AIEE

ELECTRICAL and electronic engineers are familiar with the impedance concepts of electric networks. In such networks the parameters of interest are impedances, admittances, currents, voltages, and other related attributes. As early as 1932, Foster[1] in a classic paper recognized that the electric network is a special class of a more general class of oriented graphs and geometrical circuits. More recently, Shannon,[2] Von Neumann,[3] Mason,[4,5] and others have shown that network and circuit concepts can be profitably applied to a large variety of problems, for example, the design and the analysis of switching circuits, the solution of problems in probability, the analysis of functional relationships, etc.

It is the purpose of this paper to investigate some of the properties of network which can be used to represent reliability relationships in complex equipments containing redundant elements, communication nets having many alternative communication paths, and operating systems containing alternative subsystems. In such problems each element is characterized by a single statistical parameter. A single element can be represented as a branch of a network between an input node and an output node.

If Fig. 1 represents an element of a system or an equipment, then p is the probability that the element will not fail. How is this parameter measured or determined? In the case where the element is assumed to be part of an equipment or a

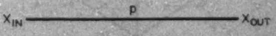


Fig. 1. An element of a system or of an equipment

Paper 58-538, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 10, 1958; made available for printing April 30, 1958.

Mr. Fred Moskowitz is with the Rome Air Development Center, Griffiss Air Force Base, N. Y.

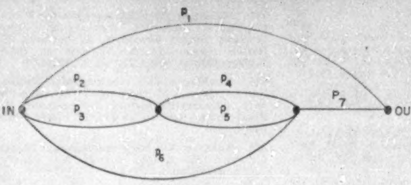


Fig. 2. A series-parallel network

system, it must be assumed that that particular element is a typical sample of a population of "identical" elements, i.e., if the average value of the population parameter is p, then individuals sampled from the population may be expected to exhibit the parameter p, except for chance variations.

If these assumptions are satisfied, the following test may be performed: Take a large sample N and subject it to a life test under specified environmental conditions. After a time t, observe the number S, that survive. The ratio of an average S to N is the survival probability p to time t.

The physical significance of the symbols in Fig. 1 in this case are: X_{in} is the number of items with survival probability p initially available, and X_{out} is the number of items expected to remain after time t. Fig. 1 then represents the linear relationship

$$X_{out}=pX_{in} \tag{1}$$

The central problem of analysis to be considered in this paper is the following: Given a system, equipmemt, or communication net, made up of a number of individual elements, each element being characterized by a reliability factor as defined by Fig. 1, find the over-all reliability factor. It will be assumed in this paper that each reliability factor is independent of all others.

In dealing with redundancy and reliability it is necessary to investigate as to how a given physical situation may be translated into an equivalent probabilistic network and then to consider various methods of solving such networks.

The advantage of the network approach in the solution of reliability and redundancy problems is the same as the advantage gained in the use of networks in the solution of electric circuit problems.

A large complex problem which cannot be solved all at once may be reduced to a series of smaller problems, each of which is amenable to solution from elementary considerations.

Canonical Form for Redundancy Networks

In principle all redundancy networks, no matter how large or complex, may be solved by summing all the possible probabilities favorable to the event that "the networks will not fail." Each additional element in the network doubles the total number of possible combinations; thus for 10 elements more than a thousand terms must be considered, while for 20 elements, the number of terms is greater than one million. This limits the simple exhaustive procedure to very small networks in practice.

In general, for n elements, the canonical equation can be written as:

$$\begin{aligned} &p_1p_2\ldots p_n+p_1q_2\ldots p_n+p_1p_2\ldots q_n+ \\ &\qquad q_1p_2\ldots p_n+\ldots+q_1q_2\ldots q_n=1 \end{aligned} \tag{2}$$

The techniques to be described permit the breaking down of complex networks into a series of elementary ones and, conversely, to compose complex networks by proper combination of simpler ones. Therefore it will be necessary to have recourse to the canonical formulation only for formal proofs of theorems.

Elementary Networks

All redundancy networks may be broken down into or built up from combinations of the following three basic elementary types: series networks, parallel networks, or bridge networks.

SERIES NETWORK

All elements in series can be replaced by a single element having a reliability factor equal to the product of the reliability factors of the elements being replaced.

For elements in series, then

$$P_{series}=\prod_{i=1}^{n} p_i=p_1p_2\ldots p_n \tag{3}$$

For the case where all the elements have identical values

$$P_{series}=p^n \tag{4}$$

PARALLEL NETWORK

A network consisting of all elements in parallel can be replaced by a single element having a reliability factor equal to one, minus the product of the unreliability factors of the elements replaced. Therefore

$$P_{parallel}=1-\prod_{i=1}^{n} q_i=1-(q_1q_2\ldots q_n) \tag{5}$$

where

$$q_i=1-p_i$$

For the case where all elements have identical reliabilities

$$P_{parallel}=1-q^n \tag{6}$$

A convenient shorthand notation to use for parallel networks is the "and/or" operator, designated in this paper by a cross, X. Thus p_1 in parallel with p_2 may be written as $p_1 \mathrm{X} p_2$ and read as "either p_1 and/or p_2." For n elements in parallel equations 5 and 6 may be written as follows

$$\begin{aligned} P_{parallel}&=p_1\mathrm{X}p_2\mathrm{X}p_3\mathrm{X}\ldots\mathrm{X}p_n \\ &=\mathop{\mathrm{X}}_{i=1}^{n} p_i=1-\prod_{i=1}^{n} q_i \end{aligned} \tag{7}$$

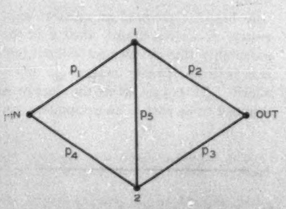


Fig. 3. A simple bridge network

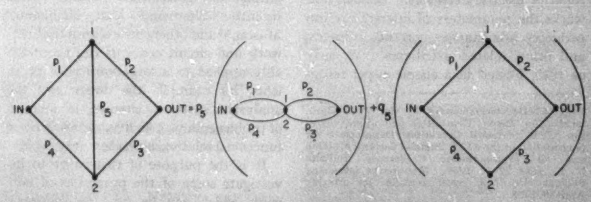


Fig. 4. The factoring theorem applied to a bridge

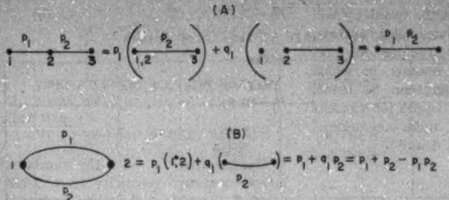


Fig. 5. Application of factoring theorem. A—Series circuit. B—Parallel circuit

SERIES-PARALLEL NETWORK

By means of two operators, the dot representing elements in series and the cross representing elements in parallel, it is possible to write the network function of any series-parallel network by inspection. The network function can then be transformed into conventional algebraic form by manipulating the function in accordance with simple algebraic rules of transformation. As an example consider the following series-parallel network consisting of seven elements. To solve the network of Fig. 2 by the canonical method would require examining 2 or 128 combinations. The network function, however, may be written directly from the diagram by applying the series-parallel principles just outlined.

The network function is

$$H(p_1,p_2,\ldots p_7)=p_1\mathrm{X}\{p_7\cdot[p_6\mathrm{X}(p_2\mathrm{X}p_3)\cdot(p_4\mathrm{X}p_5)]\} \quad (8)$$

In essence the network is solved by writing equation 8. If numerical values are assigned to the network elements, then a numerical value can be found for H by transforming equation 8 into the algebraic form in which only the operations of addition, subtraction, and multiplication exist. The transformation rules for the dot and cross operators are*

Dot operator: $x\cdot y=xy$
Cross operator: $x\mathrm{X}y=x+y-xy$

with these transformation rules, equation 8 can be expanded into standard algebraic form:

$$\begin{aligned}H=p_1+p_7[p_6+(p_2+p_3-p_2p_3)(p_4+p_5-\\ p_4p_5)-p_6(p_2+p_3-p_2p_3)(p_4+p_5-p_4p_5)]\\ -p_1p_7[p_6+(p_2+p_3-p_2p_3)(p_4+p_5-p_4p_5)-\\ p_6(p_2+p_3-p_2p_3)(p_4+p_5-p_4p_5)]\end{aligned} \quad (9)$$

Since in equation 9 conventional algebraic operations are involved, a numerical value for the over-all network reliability H may be obtained by substituting numerical values for the element reliabilities p_1, p_2, etc. Or if the value for H and values for all but one of the element reliabilities are given, it is possible to solve for the unknown element reliability.

BRIDGED NETWORKS: THE FACTORING THEOREM

Although the series-parallel technique is applicable to a large class of redundancy networks, it breaks down for networks containing cross-connected or bridging elements, such as p_5 in Fig. 3.

Bridge-type networks can be solved by applying the factoring theorem of Boolean algebra. A formal proof of this theorem is given in the Appendix.

For redundancy networks the theorem may be stated in the following form: A redundancy network function is equal to the reliability factor of any one single element of the network times the resulting network function with the terminals of the element short-circuited, plus the unreliability factor of the element times the resulting network function with the terminals of the element open-circuited.

By applying the factoring theorem to Fig. 3 the bridging element p can be removed and the bridge network transformed into the sum of two series-parallel networks which can be solved by the methods of the previous section; see Fig. 4.

The factoring theorem need not be limited to bridge-type networks, although it has been found to be most useful in reducing bridge networks to series-parallel ones by removing internal connections. As Fig. 5 shows, the elementary network transformations previously derived from other considerations may be obtained by applying the theorem to a simple series and a simple parallel circuit.

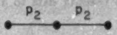
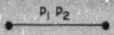
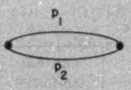
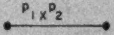
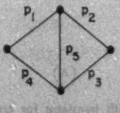
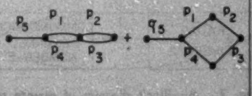
The factoring theorem does for redundancy network theory what Norton's and Thevenins's theorems do for electrical circuit theory. In each case, a complex network which is difficult to analyze as a whole can be reduced to several simpler networks. A complex problem is thus made amenable to piecemeal solution. Table I summarizes the information for elementary networks.

Problem of a Simple Bridge Redundancy Network

Consider the following hypothetical physical situation: There are two search radars, remotely located, but with overlapping search coverage patterns. Each radar is capable of a delayed video output which is relayed over individual com-

TABLE I

ELEMENTARY NETWORK TRANSFORMATIONS

	TYPE	NETWORK	EQUIVALENCE
1.	SERIES		
2.	PARALLEL		
3.	BRIDGE		 P5 P1 P2 P4 P3 + q5 P1 P2 P4 P3

---

* Note that in calculus of probability notation $x\cdot y=P(xy)$ = joint probability of x and $y=P(x)$ $P(y)$ in the event x and y are independent, and $x\mathrm{X}y=P(x+y)=P(x)+P(y)-P(xy)$.

munication channels to a central computer for further processing and assessment. Fig. 6(A) is the block diagram for this simple situation and 6(B) is the applicable redundancy circuit, if R_1 is the probability that radar 1 is operating correctly, R_2 the probability that radar 2 is operating, and C_1 and C_2 are the respective probabilities that the communication channels do not fail. It should be noted that, in this case, the computer will not be provided with correct data if, for example, radar 1 fails and, simultaneously, communications channel 2 fails, even though radar 2 and communications channel 1 are operating correctly.

To increase the reliability of the radar communications system, suppose a "translator" unit is inserted which is designed to process the delayed video output of radar 1 so that it can be multiplexed upon communications channel 2, and similarly process the output of radar 2 for multiplexing upon communications channel 1. Fig. 7(A) is the block diagram for the system including the translator and (B) is the resulting redundancy circuit, where T is the probability that the translator will not fail. If the translator fails to operate, the situation reverts to the previous case.

To illustrate the manner in which numerical computations can be made, assume the following values for element reliabilities of the radar-communications system of Figs. 6 and 7:

$R_1,\ R_2=0.8$
$C_1,\ C_2=0.9$
$T=0.8$

and let H_a be the system reliability without the bridging translator network as in Fig. 6 and H_b, the system reliability with translator included; see Fig. 7. Thus

$$\begin{aligned}H_a&=R_1C_1\mathrm{X}R_2C_2=R_1C_1+R_2C_2-R_1R_2C_1C_2\\&=(0.8)(0.9)+(0.8)(0.9)-\\&\qquad(0.8)(0.9)(0.8)(0.9)=0.922\end{aligned}$$

Note that for the values chosen, each individual radar-communications system has a reliability of 0.72. By using the two radars in parallel, the system reliability has increased to 0.922.

Adding the translator unit yields

$$\begin{aligned}H_b&=T(R_1\mathrm{X}R_2)(C_1\mathrm{X}C_2)+\bar{T}(R_1C_1\mathrm{X}R_2C_2)\\&=T(1-\bar{R}_1\bar{R}_2)(1-\bar{C}_1\bar{C}_2)+\bar{T}H_a\end{aligned}$$

where

$$\begin{aligned}&\bar{T}=1-T,\ \bar{R}_1=1-R_1,\ \text{etc.}\\H_b&=(0.8)[1-(0.2)(0.2)][1-(0.1)(0.1)]+\\&\qquad(0.2)(0.922)\\&=(0.8)(0.96)(0.99)+0.1844=0.945\end{aligned}$$

In this example, the use of the translator increased system reliability from 0.922 to 0.945, even though the translator's reliability was assumed to be only 0.8. Note also that the over-all system reliability is higher than the reliability of any of the individual components. This is not, in general, always true. If the element reliabilities are low, the system reliability may be lower than that of the individual components.

Algebra and Calculus of Redundancies

A form of Boolean algebra can be used, if properly interpreted, to simplify the analysis of redundancy networks. All variables in this algebra are real, posi-

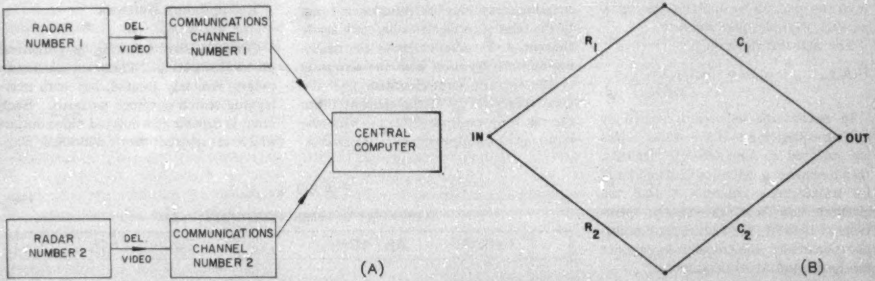


Fig. 6. A—Block diagram of two remote radars feeding data to a central computer. B—Redundancy network for radar-communications system

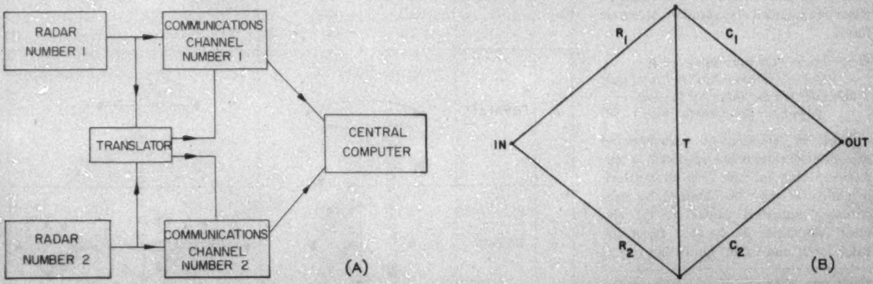


Fig. 7. A—Block diagram of remote radars with translator for cross connections. B—Redundancy network for cross-connected radar-communications system

TABLE II

BASIC OPERATIONS OF REDUNDANCY ALGEBRA

	OPERATOR	SYMBOL	OPERATION	READ AS	ORDINARY ALGEBRAIC EQUIVALENT
1	"OR"	$\oplus$	$A \oplus B$	EITHER A OR B	A+B − 2 AB
2	"AND"	•	A • B	BOTH A AND B	AB
3	"AND/OR"	X	A X B	EITHER A OR B OR BOTH	A+B − AB
4	INVERSION	BAR OVER VARIABLE	$\bar{A}$	"NOT" A	1−A

tive numbers existing in the region between zero and one inclusive, i.e., $0 \leqq A \leqq 1$. There are four basic operators: "or", "and", "and/or", and "inversion." A summary of the basic operations is given in Table II.

The symbol zero signifies absolute certainty of failure and the symbol one means absolute certainty of survival or nonfailure. The inverse of one is zero and the inverse of zero is one. Zero and one combine with redundancy variables in the following manner:

$$
\begin{aligned}
&A \cdot 0 = 0\\
&A\text{X}0 = A\\
&A \cdot 1 = A\\
&A\text{X}1 = 1
\end{aligned}
\qquad (10)
$$

The set of equations 10 are consistent with both the rules of operation given in Table II and the physical interpretation of the operations. Thus, the first equation in 10 states that the survival probability of a system consisting of an element with finite reliability in series with an element that is sure to fail is zero. Similar interpretations may be given for the other equations in the set.

The operations of division and subtraction have no significance in redundancy algebra. It is only after transformation to ordinary algebra by the rules of Table II that these additional operations become meaningful.

The rules for the operation of a variable with itself or its inverse is given in the following set of equations.

$$
\begin{aligned}
&A \oplus A = 0\\
&A \oplus \bar{A} = 1\\
&A \cdot A = A\\
&A \cdot \bar{A} = 0\\
&A\text{X}A = A\\
&A\text{X}\bar{A} = 1
\end{aligned}
\qquad (11)
$$

The proof of the dot and cross self-operations can be obtained from the following reasoning: $A \cdot A$ is the joint probability of A and A, which is the probability of A times the conditional probability of A given A, which is one. Now if $A \cdot A = A$, then $A\text{X}A = A + A - A \cdot A = A$. Similarly, $A \cdot \bar{A} = 0$, since it is impossible for the same element to survive and fail simultaneously just as it is impossible for a flipped coin to show both "heads" and "tails" at the same time. Hence, $A\text{X}\bar{A} = A + \bar{A} - A \cdot \bar{A} = 1$.

By means of these axioms and rules of algebraic manipulation the following identities may be established:

$$
\begin{aligned}
&A \cdot B \cdot C = B \cdot (A \cdot C) = C \cdot (A \cdot B)\\
&\qquad = A \cdot (B \cdot C) = (A \cdot B) \cdot C\\
&A\text{X}B\text{X}C = B\text{X}(A\text{X}C) = C\text{X}(A\text{X}B)\\
&\qquad = A\text{X}(B\text{X}C) = (A\text{X}B)\text{X}C\\
&A \cdot (B\text{X}C) = (A \cdot B)\text{X}(A \cdot C)\\
&A\text{X}(B \cdot C) = (A\text{X}B) \cdot (A\text{X}C)\\
&A\text{X}B + \bar{A} \cdot \bar{B} = 1\\
&(A \cdot B)\text{X}C - A \cdot (B\text{X}C) = \bar{A}C\\
&(A \cdot B)\text{X}A = A\\
&(A \cdot B)\text{X}\bar{A} = AB + \bar{A} = \bar{A}\text{X}B\\
&(A\text{X}B) \cdot (C\text{X}D) = AC\text{X}AD\text{X}BC\text{X}BD\\
&(A \cdot B)\text{X}(C \cdot D)\\
&\qquad = (A\text{X}C)(A\text{X}D)(B\text{X}C)(B\text{X}D)\\
&\overline{A \cdot B} = \bar{A}\text{X}\bar{B}\\
&\overline{A\text{X}B} = \bar{A} \cdot \bar{B}
\end{aligned}
$$

Other identities can be established; however, the foregoing sample demonstrates the type of equivalencies that can be obtained. The identities here are identities in the sense that numerical values can be substituted for the symbols only after the redundancy function has been transformed into ordinary algebraic form. For the redundancy diagram to be usable, it is necessary that every element in the original redundancy diagram be assigned a different symbol even if two or more elements have the same numerical value. Thus, if there are two elements in parallel, both having the same survival probabilities of 0.5, the correct redundancy function is $A\text{X}B$ not $A\text{X}A$. However, once a redundancy circuit, based upon a physically real situation, is established, it is possible to find equivalent networks. As an example of an original network see Fig. 8. An equivalent diagram is shown in Fig. 9. The equivalence is the result of the third identity in the preceding list of identities.

In Table III are listed five pairs of equivalent circuits whose equivalence can be demonstrated by writing the corresponding network functions between terminals of interest for both networks and showing that equivalent expressions can be obtained. To illustrate the method, consider the equivalent pair in item 3, Table III. For the original network

$$
\begin{aligned}
&H_{13} = AC\\
&H_{23} = AB\\
&H_{31} = BC
\end{aligned}
\qquad (12)
$$

For the equivalent network

$$
\begin{aligned}
&H_{13} = (AC)\text{X}(AB) \cdot (BC) = (AC)\text{X}(ABC)\\
&\qquad = (AC) + (ABC) - (ABC) = (AC)\\
&H_{23} = (AB)\text{X}(BC) \cdot (AC) = (AB)\text{X}(ABC)\\
&\qquad = (AB) + (ABC) - (ABC) = (AB)\\
&H_{31} = (BC)\text{X}(AC) \cdot (AB) = (BC)\text{X}(ABC)\\
&\qquad = (BC + (ABC) - (ABC) = (BC)
\end{aligned}
\qquad (13)
$$

Comparing equations 13 with equations 12 establishes the equivalence.

The other pairs may be similarly shown to be equivalent. The pair in item 4, Table III is merely that of item 3 with two additional elements added, one ele-

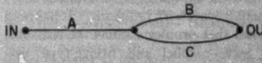


Fig. 8. Example of an original network

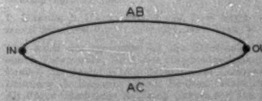
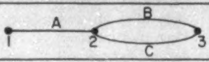
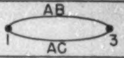
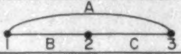
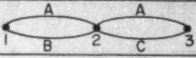
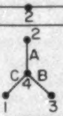
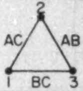
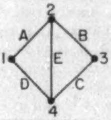
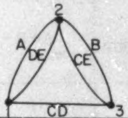
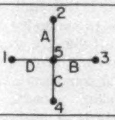
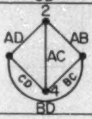


Fig. 9. Equivalent diagram

TABLE III

SOME EQUIVALENT REDUNDANCY NETWORKS

	ORIGINAL NETWORK	EQUIVALENT NETWORK
1		
2		
3		
4		
5		

ment between terminals 1 and 2 and the other between terminals 2 and 3.

Conclusions

It has been shown that if certain simplifying assumptions are made, it is possible to analyze problems reducible to probabilistic networks. The major simplification used was to assume that the elements of the network have independent statistical characteristics, i.e., the probability of failure or survival of one element of a network would not depend upon failure or survival of any other. This assumption is not fundamental. In principle, it is possible by means of more complicated network analysis, to derive network functions in which the assumption of independent failures is not made. The attempt of this study, however, was to give an approach to the solution of redundancy network problems. It has been shown that problems which are reducible to series-parallel and bridge-type networks can be solved in a simple systematic manner.

Appendix. The Factoring Theorem

Given an n element redundancy network connected in any manner whatsoever, having element reliabilities $p_1, p_2, \ldots p_n$, and network function with respect to any two given terminals, $H(p_1,p_2,\ldots p_n)$,

$$H(p_1,p_2,\ldots p_n) = p_k[H(p_1,p_2,\ldots p_n)]_{p_k=1} + \bar{p}_k[H(p_1,p_2,\ldots p_n)]_{p_k=0} \quad (14)$$

where p_k is the reliability of the kth element.

Proof: If $H(p_1,p_2,\ldots p_n)$ is a function representing a real network, then it can be expressed in canonical form. When written in canonical form, note that every term must contain either p_k or $\bar{p}_k$ but not both. Let H_a be the sum of those terms in the canonical equation which contain only p_k, and H_b the sum of those terms which contain only $\bar{p}_k$. Since H_a contains only terms with p_k, p_k may be factored out, giving

$$H_a = p_k H_s \quad (15)$$

Similarly,

$$H_b = \bar{p}_k H_0 \quad (16)$$

But

$$H(p_1,p_2,\ldots p_n) = H_a + H_b = p_k H_s + \bar{p}_k H_0 \quad (17)$$

Now, neither H_s nor H_0 contain either p_k or $\bar{p}_k$; therefore, setting $p_k=1$ for both sides of equation 17 yields

$$[H(p_1,p_2,\ldots p_n)]_{p_k=1} = H_s \quad (18)$$

since for $p_k=1$, $\bar{p}_k=0$.

Similarly, setting $p_k=0$, for which $\bar{p}_k=1$, yields

$$[H(p_1,p_2,\ldots p_n)]_{p_k=0} = H_0 \quad (19)$$

Combining equations 18 and 19 with 17 gives

$$H(p_1,p_2,\ldots p_n) = p_k[H(p_1,p_2,\ldots p_n)]_{p_k=1} + \bar{p}_k[H(p_1,p_2,\ldots p_n)]_{p_k=0} \quad (20)$$

which proves the theorem.

References

1. GEOMETRICAL CIRCUITS OF ELECTRICAL NETWORKS, Ronald M. Foster. *AIEE Transactions*, vol. 51, June 1932, pp. 309–17.

2. RELIABLE CIRCUITS USING LESS RELIABLE RELAYS, C. Shannon, F. S. Moore. *Journal*, Franklin Institute, Philadelphia, Pa., Sept. and Oct. 1956.

3. PROBABILISTIC LOGICS (book), J. Von Neumann. California Institute of Technology, Pasadena, Calif., 1952.

4. FEEDBACK THEORY, SOME PROPERTIES OF SIGNAL FLOW GRAPHS, S. J. Mason. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Sept. 1953, pp. 1144–56.

5. FEEDBACK THEORY, FURTHER PROPERTIES OF SIGNAL FLOW GRAPHS, S. J. Mason. *Ibid.*, vol. 44, July 1956, pp. 920–26.

6. SIGNAL FLOW GRAPHS AND RANDOM SIGNALS, W. H. Huggins. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, No. 1, Jan. 1957, pp. 74–86.

7. SOME RELIABILITY ASPECTS OF SYSTEMS DESIGN, F. Moskowitz, J. B. McLean. *Transactions*, Professional Group on Reliability and Quality Control, Institute of Radio Engineers, New York, N. Y., Sept. 1956, pp. 7–35.

8. MATRIX ANALYSIS OF ORIENTED GRAPHS WITH IRREDUCIBLE FEEDBACK LOOPS, J. K. Percus. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-2, June 1955, pp. 117–27.

9. PROPERTIES OF TELECOMMUNICATIONS NETWORKS, Z. Prihar. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, July 1956, pp. 927–33.

10. ORDER OF COMPLEXITY AND MINIMAL STRUCTURES IN NETWORK ANALYSIS, F. M. Reza. *Technical Report*, Massachusetts Institute of Technology, Cambridge, Mass.

Operations Research in Communications

B. HARRIS
NONMEMBER AIEE

A. HAUPTSCHEIN
NONMEMBER AIEE

L. S. SCHWARTZ
MEMBER AIEE

OPERATIONS RESEARCH can be of value to the communication engineer by enabling him to set up valid criteria by which to judge the relative advantages of communication systems in specified applications. While in many cases insufficient data are available for an exact analysis or optimization, this approach is superior to the classical method of choosing arbitrary evaluation parameters which may be meaningless in the proposed application. On the other hand care must be taken in the interpretation of the results of an operations research procedure. In the classical approach the system is designed to meet some given specifications and is not optimum in any particular sense. In an operations research approach criteria of performance and effort, plus constraints, are chosen, and the system is optimized with respect to these quantities. However, because of necessary simplifications in both the criteria and the constraints the results are only pseudo optimum with respect to the desired applications. In effect, the ambiguities involved in choosing the specifications of a classical approach may have been exchanged for the ambiguities involved in choosing the criteria and the constraints in an operations research approach. Nevertheless, an operations research approach is desirable because it helps prevent unintentional ambiguities from entering the problem by directly focusing attention on the application as well as the system.

Historical Background

Operations research is as old as science and management. Its development as a recognized form of research, however, was in response to the military needs of World War II. In this activity scientists were attached to the staffs of military organizations for the purpose of giving scientific advice on those matters which were not handled by the service technical establishments. The staffs of the military provided the scientist with the operational outlook and data, and the scientists applied the scientific methods of analysis to these data and were thus able to give useful advice. The scientist used as data the material to be found in an operation room, e.g., signals, track charts, combat reports, meteorological information, etc. Experience showed that the analysis of operations in a more scientific manner than is done normally by service specialists or officers was of utmost value. This was especially the case with those aspects of operations into which probability and statistical considerations enter.

One important application of operations research during World War II dealt with the systematic evaluation of enemy countermeasures.[1] Without this objective evaluation the tendency was to credit the enemy with an effective countermeasure before it was actually developed. For example, consider the question of installing early-warning radar on submarines. A radar set which could detect airplanes farther off than lookouts would enable the submarine to dive in time to escape detection. On the other hand, many submarine skippers were convinced that the Japanese planes were equipped with radar receivers and were thus detecting the submarines by means of the radar signals. This view was based on the fact that more Japanese planes came into view when the early-warning radar was on than when it was off.

The radar should "see" more airplanes because its radius of detection is greater. In fact, if the planes are not exploiting the radar signals, the ratio of the numbers of aircraft seen with and without the radar will equal the ratio of the corresponding radii of detection. Since the data confirmed this, early-warning radar was used on U.S. submarines. As a result, in spite of the fact that the first early-warning radar improved the detection range by only 40%, U. S. submarines were caught on the surface in less than 5% of the attacks. The Japanese, on the other hand, did not install radar and U.S. planes enjoyed a 50% chance of surface attack.

Since World War II, operations research has been applied increasingly to the solution of problems of both military and commercial interest such as traffic delays at toll booths,[2] New York's subway fare structure,[3] combat stress in Korea,[4] supermarkets,[5] and the hand setting of type in the printing industry.[6] One example of a commercial application is in the realm of hotel management. At any given time a hotel has a number of guests in residence and a number of advance reservations. How many additional reservations should the booking clerk accept? If he turns away all reservations when the hotel is filled on paper, he will rarely have a filled hotel, for on the average about one out of five reservations never show up. On the other hand, he cannot always accept all reservations, for occasionally he will be called on to house more people than the hotel has room for. He must somehow balance the gain in having a large number of rooms rented against the loss in good will of occasionally failing someone who has been given a reservation. As soon as the hotel manager has decided how much such a failure costs the hotel, the other factors in the analysis involve only straightforward probability analysis. The computation of the mean-stay of a guest and its relation to the check-out rate of guests already present, the curve for hotel reservation cancellations, the dependence of fluctuations on time of week, and so on can be reduced to a fairly simple set of rules for the booking clerk to follow in making his day-by-day decisions.

Principles of System Evaluation

In a fully developed operations research process four fundamental steps or actions are necessary: (1) choice of criterion of performance, (2) choice of measure of effort, (3) more or less explicit setting up of a model, and (4) optimization. Of these four actions the selections of a criterion of performance and a measure of effort are the most important. These are the crucial steps in operations research because the concept of value and the standard by which the relative importance of the results can be judged is involved. In the strictest sense those quantities which suitably measure the performance and effort of a system cannot be precisely defined unless the exact employment of the system and the utility evaluation of the decisions are both known. In most instances, therefore, the high-level problem of over-all optimization is extremely

---

Paper 58-575, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28–30, 1958, and re-presented at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted January 3, 1958; made available for printing April 21, 1958.

B. Harris, A. Hauptschein, and L. S. Schwartz are all with New York University, New York, N. Y.

This work was sponsored by the Electronics Research Directorate, Air Force Cambridge Research Center, Air Research and Development Command, Laurence G. Hanscom Field, Bedford, Mass., the Air Force Cambridge Research Center and the Air Research and Development Command, Baltimore, Md., under Contract no. *AF 19(604)-1049*.

difficult and suboptimization of the system components is sought instead.

Frequently one of the most important contributions of an operations research approach lies in determining the proper criterion of performance and measure of effort. That the criteria chosen greatly affect the conclusions can be seen by considering the problems the British had in World War II in deciding whether or not to install antiaircraft guns on merchant vessels.[1] The installations of antiaircraft guns on merchant vessels were not only expensive, but in many cases the guns were needed elsewhere. Moreover, the guns were relatively ineffective so far as an aggressive criterion was concerned because, with inexperienced crews, planes were shot down in only 4% of the attacks. This seemed to indicate that the guns should not be installed. However, as further data were taken, it became apparent that the proper criterion was not the number of planes destroyed but rather the protection the gun afforded the ship. Analysis showed that the accuracy of the attacking plane was reduced by the gun fire, so that the chance of the ship being sunk was reduced in half. This greatly favored the installation of the guns and decided the question.

However, the defensive criterion of protecting the ship is not really the criterion used in this low-level problem. If it were, the conclusion would be not to install the guns but to keep the ships in port. Criteria such as this, which must be hedged to present absurd results, should be avoided. Instead the high-level approach should be considered, namely, whether the installation of guns on merchant ships would aid the conduct of the war, and the low-level criteria chosen so as to be consistent with the high-level criteria. A suitable criterion in this case would combine the cost of installations against the increased number of successful trips and the materials lost when ships were sunk. That operation which minimized a total cost such as this would be optimum.

The third and fourth actions in an operational research analysis deal with the procedural steps involved in evaluating given situations once the figures of merit and the measures of effort have been set up. It has been unfortunately true that whereas the choice of criteria and measures of effort are central to the operations research problem, almost all the effort thus far has been devoted to the purely procedural phases involved in actions 3 and 4. An important reason is that technical workers are of the conviction that since philosophizing about criteria involves value judgments, it is not science and, therefore, not quite respectable. As long as value judgments remain not-science, however, operations research cannot be science either.

Application to Communication Systems

In the remainder of this paper an attempt is made to suggest a method of obtaining a criterion for evaluating the performance of communication systems, a criterion whose use, it is hoped, will result in more than suboptimization. Perhaps, by means of this criterion it will be possible to evaluate system performance quantitatively in an economic sense. In evaluating communication systems one must realize that the purpose of a communication link is to transmit correct messages. Unless this restriction is made one has the philosophical problem of handling situations where a wrong message may be better than the right one, and where the worth of a communication changes with hindsight. It should be noted that an operations research approach to the communication problem is based on evaluating the future behavior and worth of a system in terms of present knowledge concerning the proposed applications.

For purposes of analyzing the over-all operational situation it is convenient and conventional to consider the system as two sequential actions: that of transmitting information, and that of acting on the received information. When the behavior of the action link is unknown, the communication link must be characterized by the conditional probabilities relating all the possible perturbations of one symbol into another. Only by including this information can an over-all performance measure be evaluated for particular rather than average messages, as for example, by specifying both the mean and the confidence limits on the expected utility. This is of concern when a particular communication may be of extreme importance and certain message perturbations must be avoided or be under control.

As in many analogous instances, however, one is willing to forego accuracy for simplicity. In this instance, since the short-term employment of the system and the utility evaluations of the resulting actions are usually only vaguely known, foregoing accuracy in the specifications of the communication link may result in only a small change in the over-all accuracy of the performance measure. Hence concern is conventionally limited to average message reliability parameters and average employment characteristics. This approach permits an integrated parameter to be substituted for the many conditional probabilities otherwise involved. Which integrated parameter is actually indicative of the over-all reliability still depends, of course, on the worth of the behavior based on the received signal.

In conformity with this viewpoint two costs may be considered: (1) the mean communication cost or measure of effort associated with the transfer of information, and (2) the mean risk cost associated with the resulting decisions.

Communication cost is determined by the design and operation of the equipment and thus indirectly by the power, bandwidth, and time requirements. It also depends on the expected operational conditions of message density, the method of modulation, and the system of information coding. Risk cost, on the other hand, is a function of the receiver's ability to distinguish signal from noise or the transmitter's intent from the interference. In general, increases in power, bandwidth, and time decrease the risk cost and increase the communication cost.

In the optimization problem the actual communication cost is evaluated by knowing the fixed costs as well as the operational environment in which the system performs. The risk cost is determined from knowledge of the manner in which the messages are perturbed by interference and the cost of resulting wrong decisions. The total cost may then be considered the sum of the communication cost and the risk cost and the best system operation determined by standard minimization procedures. In this manner, minimum costs can be determined for other systems and comparisons made among them. The nature and interrelation of these two costs is now considered.

Communication Cost[7]

In establishing communication cost relationships, one must include initial construction and installation costs, maintenance costs, and operational costs. Initial and maintenance costs are fixed costs and do not vary significantly with the operational situation in which they are used. On the other hand, operational costs, as defined here, are functions of the operational conditions.

A necessary first step in a cost formulation is to make a statement of preferences. In terms of cost this means that relative weightings are assigned to the basic parameters of the communication problem.

Relative weightings enter because, depending on the operational situation, the basic parameters may have different relative importance. In communication systems three parameters are of fundamental importance and may be considered as basic parameters. These are signal power, signal bandwidth, and operational time. Time, as used in this context, has two aspects. One aspect refers to the transmission time of the messages over the channel. This aspect is a function of the information, the information rate, and the coding delay. The other aspect refers to the waiting time and is a function of the percentage of full loading of the facilities and the number of communication channels. This latter aspect is concerned with traffic density and problems of congestion such as arise in telephone operation and in aircraft control at large airports. Transmission of information thus results in an information time delay, while loading results in waiting time delay. The sum of these time delays is called the operational time.

Initial cost is a function of power, because power output determines the size and weight of a transmitter; it is a function of bandwidth, because bandwidth determines the size, character, and number of components; and it is a function of time, because time, when utilized in a trade-off with power and bandwidth, implies the use of coders, decoders, and noise filters. Maintenance cost is a function of all three basic parameters, because they determine equipment complexity. Operational cost, however, is determined by the relative importance of the basic parameters in the operational situation.

As an example consider the problem of supplying a walkie-talkie to a combat patrol. There are initial costs of construction, installation costs of delivery, and maintenance costs. In actual use the equipment has an operational cost due to such factors as its weight and bulk (i.e. a rifle in place of the walkie-talkie might be preferred), the possibility of the signal being detected by enemy forces, and the possibility of its interfering with other communication equipment. These factors are reflected in the signal bandwidth and power. Indeed, to keep these costs low the bandwidth and power output of a walkie-talkie is fairly low although the operational time, i.e. the number of times a message is repeated to assure transmission, may be fairly high.

RISK COST

Risk cost is concerned with the reliability of system operation and also is a function of the basic communication parameters of power, bandwidth, and time. Risk cost enters because in the reception of information over a noisy communication channel, decisions must be made according to some criterion. These decisions can be made according to any one of several criteria, but it can be shown[8] that there is a best decision rule, in the sense of minimizing the average risk, once the statistics of the noise and signal are known. Errors will be made, however, even under the best decision rule.

It is these errors in decision which produce the risk cost, since they can have bad consequences in the operational sense. For example, errors due to noise in a control link of an antiaircraft system might cause the incorrect aiming of a gun which is attempting to hit a bomber. The reward for obtaining correct control information is the destruction of the bomber, while the penalty for incorrect control information is serious damage to ground installations. Since the reward for a correct decision is, in general, in different units than that of the penalty for an incorrect decision, some universal performance measure is desired. Such a measure may be obtained by assigning a cost or utility value to each decision, i.e. to every combination of system input and system output. As mentioned previously, such assignment of cost to each decision is usually not practical. Hence it is desired to determine the cost of average performance. This may be done by substituting an integrated reliability parameter for the complex of associated conditional probabilities. Several possible reliability parameters which may be used are: rate, per-unit equivocation, accuracy, error probability, average error per bit of information, and generalized error probability. In some instances, of course, none of these parameters may describe the situation. As an extreme case, for example, the entire asset of a walkie-talkie may be in the increased morale of the patrol and errors in transmission may have, within large limits, negligible effect.

One convenient bound on the performance of communication systems is that presented by Shannon.[9] The significance of this result is that if infinite coding delays are permissible, and equipment complexity is of no consideration, error-free messages may be sent even through a noisy communication link. The only effect of the noise is to limit the maximum information rate which can be handled by the channel. Since in general the coding delay and equipment complexities are important, actual transmissions cannot in general be error-free. One interesting exception to this rule is an information feedback system where there is a noise-free feedback path of at least the same capacity as the noisy direct channel.[10]

In the general information feedback system the receiver suitably confirms the received information and the transmitter decides whether or not the receiver has properly interpreted the message. From the viewpoint of reliability, information feedback can therefore accomplish a great deal, particularly if the feedback channel is error-free. The term "information feedback" is used to distinguish this form of feedback from decision feedback[11] where the receiver decides whether or not the received message can be properly interpreted and then reports its decision to the sender.

A simple information feedback system is an iterative discarding system in which the feedback process is used to improve the confirmation denial information as well as the main message. For example, the system can operate as follows. If the feedback message indicates that an incorrect direct channel message was received, the sender transmits an erasure signal followed by the correct message. If it is indicated that an erasure was incorrectly received as a message, the sender transmits two erasures to erase both this incorrect message and the preceding incorrect message. When a message is received as an erasure, the sender merely transmits the preceding message and the intended message. If the feedback is error-free and totally distinctive, all errors will eventually be corrected and the final message will be error-free. The unique feature of this system is that error-free transmission can be obtained without infinite delay, whereas this result may never occur in unidirectional systems. The main penalty for this error-free transmission lies in the fact that the discarding process discards large amounts of information at low signal-to-noise ratios. Thus there exists a threshold signal-to-noise ratio below which the information rate becomes zero.

In this nontruncated feedback system the final message is error-free, but there exists the possibility of large storage delays before a message is finally accepted, and such delays may exceed the memory capacity of the equipment. Hence in a practical communication channel, there will be some maximum number of repeats which cannot be exceeded, so that the deterioration in system performance due to truncation of the feedback process is of interest.[12] As expected, the primary effect of truncation is, of course, to prevent error-free transmission.

Except for special situations, which do not occur in practice, a message will not be error-free, and thus some risk parameter must be chosen to describe the worth of the system to the receiver. Several such parameters are described in the following.

(*a*). *Rate, Per-Unit Equivocation, and Accuracy*

For example, consider a communication channel which is being used to transmit the results of a chance situation to a gambler before these results become common knowledge, so that the gambler may place bets at the original fair odds.[13] By placing suitable bets the gambler may, with probability one, insure that his capital will increase at an exponential rate, the exponent being exactly the Shannon information rate. Since the information rate is given by

$$R=H(x)-H_y(x)$$

where $H(x)$ is the information of the informer and $H_y(x)$ the information destroyed by the noise in the channel, a suitable measure of the usefulness of the channel to the receiver is the information rate or the ratio of $H_y(x)$ to $H(x)$, i.e. the per-unit equivocation. In fact, any communication system which utilizes all a posteriori probabilities, and whose purpose is to increase the chance of winning in a situation where the loss, if incorrect, and the gain, if correct, is proportional to the amount invested, can be characterized by per-unit equivocation as a reliability parameter.

Another form of a risk measure based on rate is accuracy. Accuracy was originally introduced by Laemmel[14] and is defined as the information rate divided by the error entropy in bits per symbol, i.e.,

$$A=\frac{R}{H_z(y)}=\frac{H(y)-H_z(y)}{H_z(y)}$$

This ratio can be shown to behave similar to signal-to-noise ratio for certain systems, varying between 0 and ∞ as reliability increases.

(*b*). *Probability of Error*

Another reliability parameter which is widely used is probability of error. The general class of situations for which probability of error is applicable is that for which the mathematical model can be simplified to the extent that only one conditional probability is necessary to specify the system. That is, all kinds of errors are considered equally bad, or a uniform risk cost is associated with all wrong decisions. Then, while the worth of the system may be a complicated function of the error probability, it is nevertheless true that specifying one specifies the other. This concept may be illustrated by an antiaircraft gun example. If a significant hit probability exists only when the gun receives correct control or direction information, then all control errors will result in missing the bomber, thus incurring equal risk costs.

For an unidirectional system the average error probability may be obtained by averaging over the error probabilities associated with each possible input signal, For a bidirectional or feedback system the concept of a generalized error probability, discussed in section *d*, may be necessary.

(*c*). *Average Error Per Bit of Information Transferred*

The average error per bit of information is a risk parameter that attempts to combine both rate and error probability concepts. By definition

$$\xi=\frac{P_e}{R}=\frac{P_e}{H(x)(1-E)}$$

where E is the per-unit equivocation, $H_y(x)/H(x)$, and P_e is the probability of error. This risk measure also shows the desired behavior with reliability, i.e., for reliable systems P_e becomes small, and ξ approaches zero. Conversely, ξ becomes large for a poor system. The measure is easy to calculate provided R is given. However, since R is a nonlinear function of P_e, the knowledge of P_e, except in certain isolated cases, permits the rate of information to be only approximated. Also, if R and ξ are given, it is not possible to state the exact amount of information lost, but bounds on lost information for specific systems can be obtained. For example, bounds on equivocation and error entropy can be determined for b-level pulse-amplitude-modulated systems having fixed P_e.[11]

(*d*). *Generalized Error Probability*

The direct extension of the error probability concept to communication feedback systems,[15] whether they be information, decision, or compound feedback systems, cannot be made because, in general, there is a nonconservation of message length and thus no one-to-one correspondence exists between the input and output symbols comprising the message. As an example, consider the information feedback system previously described where 0 and 1 are the input message symbols and φ the erasure symbol. To send a simple message, such as (0 1 0), the actual transmission, in the case of noisy feedback, could be

Direct channel message transmitted........................0 φ 0 1 0

Direct channel message received and repeated in the feedback channel........................0 0 0 1 0

Feedback message received......1 φ 0 1 0

In this case, corresponding to the transmission of a 3-digit message, not only are five digits received and accepted, but one error in transmission has occurred in the direct channel and two errors in the feedback channel. On the other hand, if the actual transmission is

Direct channel message transmitted. .0 1 0

Direct channel message received and repeated in the feedback channel.....0 φ 0

Feedback message received.........0 1 0

the received message contains only one unerased digit. To complicate matters further the actual transmission could be such that errors in transmission are made, but the received message is correct after the erasure process has been applied.

In order to account for these possibilities the error probability concept must be generalized. One such generalization is based on considering the entire message as being right or wrong. Consider, for example, an unidirectional system where there is no intersymbol constraints and a one-to-one correspondence between the input and output symbols. Then the probability of a message of M symbols being correct is

$$Q_{\text{message}}=(1-P_{\text{symbol}})^M$$

where P_{symbol} is the error probability per symbol. This indicates an exponential rate of decrease, with increasing message length, in the probability of a message being correct. It suggests that for long messages P_{symbol}, defined as

$$P_{\text{symbol}}=\lim_{M\to\infty}\{1-(Q_{\text{message}})^{1/M}\}$$

be considered as a generalized error probability per symbol. In comparing systems, whether they are unidirectional or bidirectional, the system having the smallest generalized error probability per symbol will, for long messages, have the greatest probability of correct transmission.

This presumes, of course, that the probability of a message of M symbols being correct, for the systems under consideration, decreases exponentially with increasing message length. In this event Q_{message} has the form

$$Q_{\text{message}}=f(M)e^{-\alpha M}$$

where the error exponent α is not zero and $f(M)$ is always less than M^n for some finite n. The evaluation of P_{symbol} then constitutes the evaluation of α, larger values of P_{symbol} corresponding to smaller values of α. Specifically

$$-\alpha = \lim_{M\to\infty}\left\{\frac{1}{M}\ln Q_{\text{message}}\right\}$$

or

$$-\alpha = \ln(1-P_{\text{symbol}})$$

The actual evaluation of the generalized error probability can be done by considering an input message of length M. The probability of such a message being received correctly is then determined, and the corresponding average probability of receiving messages of length M correctly is found by averaging over all messages of such length. The process is then repeated for longer messages and the asymptotic value of $1-(Q_{\text{message}})^{1/M}$ determined. This then corresponds to the generalized error probability per symbol.

It should be noted that the choice of a risk measure in some instances may not be obvious, and any one of several might give significant comparative results in system evaluation. The foregoing remarks are presented to point out some of the various measures that do exist, and for which in some operational environments one form might be clearly more applicable than another. The advantages or disadvantages of using one risk measure rather than another lies solely in its suitability to describe a situation more accurately.

COMBINED COMMUNICATION AND RISK COST

In obtaining a complete evaluation of a system, both communication and risk costs play a part. The total cost for the system is the sum of the two costs, accounting for the salient factors in a communication situation. Thus, for a particular system, the average total cost is

$$C=\int_L\int_N P(N;L)[C_r(N;L)+C_c(N;L)]\,dN\,dL$$

where $P(N;L)$ is the probability of the noise or interference being N when the operational situation, such as the message density, is L, C_r is the corresponding risk cost, and C_c the communication cost. In general the evaluation of the risk cost is itself an averaging process over all possible message alterations that can occur in transmission. Thus, for example, when a discrete set of messages is employed, and the error probability is a significant risk measure,

$$\begin{aligned}C_r(N;L)&=C_r(P_e;L)\\&=C_r\Big(\sum_m P(m)P_e(m,N);L\Big)\end{aligned}$$

where the error probability P_e must be evaluated by averaging the probability of a particular message m being transmitted, $P(m)$, against the probability of that message being received in error, $P_e(m,N)$, when the interference is N. Since in general the risk cost is reduced if the error probability is reduced, a method which so reduces the error probability at a "small" increase in communication cost is considered valuable as a suboptimization. Actual optimization must, of course, evaluate the performance in terms of the total cost. This requires that the total cost be written in terms of perhaps power, bandwidth, and time as basic parameters. By setting

$$\frac{\partial C}{\partial P}=0,\qquad \frac{\partial C}{\partial W}=0,\qquad \frac{\partial C}{\partial T}=0,$$

the actual power, bandwidth, and time at which optimum performance results can be obtained. The system specifies the error probability P_e, but the cost relationships must be determined by collection of sufficient pertinent data. This process is thus one which rightly falls into the realm of operations research, where the necessary weighting of the merit parameters must be obtained from a study of past and current operations and introduced into the theoretical model of the situation. Once the model has been found to be a valid representation, the values of the parameters may be altered in order to predict the effects of new courses of action.

References

1. **Methods of Operations Research**, P. M. Morse, G. E. Kimball. Published jointly by The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley & Sons, Inc., New York, N. Y., 1951.

2. **Operations Research for Management** (book, vol. 2), J. F. McCloskey, J. M. Coppinger, editors. The John Hopkins Press, Baltimore, Md., 1956. "Traffic Delays at Toll Booths," L. C. Edie (originally published in *Operations Research*, Baltimore, Md., vol. 2, no. 2, 1954, pp. 107–38).

3. *Ibid.* "Revising New York's Subway Fare Structure," W. S. Vickrey (originally published in *Operations Research*, vol. 2, no. 1, 1955, pp. 38–68).

4. *Ibid.* "A Study of Combat Stress in Korea," S. S. Davis, pp. 250–63.

5. **Operations Research for Management** (book), J. F. McCloskey, F. N. Trefethen, editors. The John Hopkins Press, Baltimore, Md., 1954. "Opportunities for Operations Research in Supermarkets," S. T. R. Abt, pp. 289–304.

6. *Ibid.* "Operations Research in the Printing Industry," R. H. Roy, pp. 321–41.

7. **Merit Criteria for Communication Systems**, A. Hauptschein, L. S. Schwartz. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., vol. 3, no. 1, 1957, p. 52.

8. **Modern Statistical Approaches to Reception in Communication Theory**, D. VanMeter, D. Middleton. *Ibid.*, vol. 4, 1954, p. 119.

9. **A Mathematical Theory of Communication**, C. Shannon. *Monograph B-1598*, Bell Telephone System, New York, N. Y., 1948.

10. **Theory of Information Feedback System**, S. S. L. Chang. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 2, no. 3, 1956, p. 29.

11. **Supplementary Notes on Evaluation Theory for Communication Systems**. *Fourth Quarterly Report*, Contract no. *AF 19(604)-1049*, Armed Services Technical Information Agency, Dayton, Ohio, Mar. 1955.

12. **Minimum Cost Decision-Feedback Systems for Detecting Signals Perturbed by Additive Gaussian Noise**, B. Harris, A. Hauptschein, L. S. Schwartz. *Operations Research*, vol. 5, no. 5, 1957, p. 680.

13. **A New Interpretation of Information Rate**, J. L. Kelly. *Bell System Technical Journal*, New York, N. Y., vol. 35, no. 4, 1956, p. 917.

14. **Characteristics of Communication Systems**, A. E. Laemmel. "Report on Phase II of the Study of Data Transmission and Appendices 1–6," *Report R-233-50*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., no. 178, June 1951.

15. **Binary Communication Feedback Systems**, B. Harris, A. Hauptschein, K. C. Morgan L. S. Schwartz. *AIEE Transactions* vol. 77, pt. I, 1958 (*Paper No. 58-1233*).

◆

Discussion

R. Krzyczkowski (Westinghouse Electric Corporation, Baltimore, Md.): The authors deserve to be congratulated for their skillful introduction to operations research. However, I do not think that they have done justice to the subject for it is simply impossible in the context of one brief paper. The art of operations research (or is it science?) is very wide and grows with increasing acceleration. Thus, all references quoted in the paper date between 1954 and 1956. Recently, I completed a 2-week operations research course at the Johns Hopkins University; they considered their own texts, issued in 1956, as obsolete and gave us a new book just published.

Operations research has very many applications in communications not solely of economic evaluation nature. In our organization, teams engaged in over-all military systems design comprise not only engineers, mathematicians, and economists, but also military experts and human engineers, and we think that a lawyer and a sociologist could be usefully and fully occupied. Authors seem to be under the influence of the somewhat dessicated mathematical Massachusetts Institute of Technology school of operations research. We, probably under the influence of Johns Hopkins, tend to think that the wide, Case School approach may be more appropriate for telecommunication systems, that is, for man-machine systems.

However, to confine myself to the paper under discussion, the authors' choice of basic parameters of power, bandwidth, and time as main factors in the cost function is very interesting and theoretically gives opportunity for very elegant mathematical treatment. Certainly it establishes judgment criteria. Unless authors can unveil

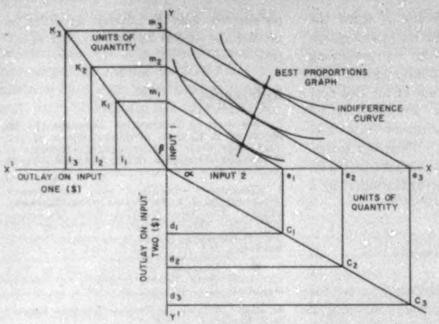


Fig. 1. Construction for finding the least outlay combination

practical results obtained, though, I am somewhat sceptical as to the usefulness of this tool for actual designs of systems or equipment.

First, if parameters are restricted to three (power, bandwidth, time) or even four (adding reliability), equipments cost compared must be of the same class. They must be all radio, for instance, and not cable versus radio as parameters immediately lose their significance. Moreover, if we consider radio, say, equipment should be of the same type, e.g. troposcopic scatter, work in the same frequency band, and in general be quite similar.

Similar work is no doubt carried out in several organizations. At Westinghouse, I made a comparison of competing scatter systems using the same principle. However, my parameters were less "theoretical." I used power transmitted, amount of diversification, diameter of antenna, and receivers' noise figure. These four were most significant from cost point of view.

In a recent paper[1] I used the same method and actually gave dollar cost figures for 18 different cable systems. Four parameters are involved: size of cables; number of cables; facilities offered; and transmission techniques used. I chose these after a long struggle with ideas like dollars per decibel or dollars per kc. In my experience they were just not helpful as soon as different systems had to be compared.

The cost equation is not the only approach to function minimization. Some 3 years ago I applied the indifference curve/price line diagram technique, as evolved by professional economists such as Stigler, Samulson, and Boulding here, or Hicks in England, to communication systems design purposes.[2] The principle of such a construction is shown in Fig. 1. Input *1* is, say, power; input *2*, bandwidth. Each indifference curve represents a series of systems which are equally good, or bad, systems where trading of some power results in gain in bandwidth. The higher the curve, the better the system. Straight lines are price lines; they show how much you can afford to spend on the system. Obviously you tend to choose the highest curve, and the tangency point gives you the best combination of the two inputs. The diagram is 2-dimensional but algebraically one can treat any number of parameters; Lagrange multipliers are most useful in this calculation. In my experience, this method is very useful, especially for problems concerning growing systems.

REFERENCES

1. DETERMINATION OF THE LEAST FIRST COST OF A CABLE-CARRIER SYSTEM, **Roman Krzyczkowski.** *AIEE Transactions*, vol. 77, pt. I, Nov. 1958, pp. 778–82.

2. AN ECONOMIC ANALYSIS OF EQUIPMENT OUTLAYS IN CABLE CARRIER SYSTEMS. *A.T.E. Journal*, Liverpool, England, vol. 11, no. 4, Oct. 1955, p. 249.

On Coding for the Binary Symmetric Channel

A. B. FONTAINE
ASSOCIATE MEMBER AIEE

W. W. PETERSON
ASSOCIATE MEMBER AIEE

QUESTIONS concerning coding for the binary symmetric channel might be divided into three categories:

1. For a specified error probability and with no restriction on the length of a code, what is the maximum rate at which information can be transmitted, i.e., what is the channel capacity?

2. For a specified code word length n and a specified rate less than the channel capacity, what is the error probability for the optimum code?

3. How can optimum codes, or at least good codes with a given rate and word length be constructed?

The first question was answered by Shannon.[1] The second question has been answered to a large extent by Slepian[2] and Elias[3–5] and in an unpublished paper by Elias. Slepian has compiled a table of optimum group alphabets for codes of small n. Elias has given upper and lower bounds on the probability of error for optimum alphabets for the binary symmetric channel. These upper and lower bounds are very close over a wide range of rates for large n, but for values of n in the order of the upper limit in Slepian's table, the upper and lower bounds may differ by several orders of magnitude. Codes in this range are now being considered for use in practical systems.

The third question, that of finding optimum codes, or at least good codes, is the most difficult. Slepian's table of optimum group alphabets is a good start. To it can be added Golay's "lossless code"[6] with 23 bits of which 12 are information bits. The Hamming codes[7] are optimum and codes described by Reed[8] and MacDonald[9] are at least good codes; and among all three types are arbitrarily long codes. Hamming's codes are high-rate codes—as n approaches infinity, the ratio of the number of check bits to the number of information bits approaches zero. Reed's and MacDonald's codes are both low-rate coding systems—as n approaches infinity, the ratio of the number of information bits to the number of check bits approaches zero. There are, however, virtually no results for the middle range of rate except for small values of n.

The principal results in this paper deal with medium length codes in this middle range of rate. Elias' bounds on error probability involve some approximations which are made for the sole purpose of obtaining an end formula from which cal-

---

Paper 58-923, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 17, 1958; made available for printing April 17, 1958.

A. B. FONTAINE is with the International Business Machines Corporation, Yorktown Heights, N. Y., and W. W. PETERSON is with the University of Florida, Gainesville, Fla.

00000	01011	10110	11101	**Code Points**
00001	01010	10111	11100	**Other Received Sequences**
00010	01001	10100	11111	
00100	01111	10010	11001	
01000	00011	11110	10101	
10000	11011	00110	01101	
01100	00111	11010	10001	
00101	01110	10011	11000	

Fig. 1. Example of the standard array of a binary code for $n=5$ and $k=2$

culations can be made readily. Calculations in which some of these approximations were not used were made on the IBM (International Business Machines Corporation) *704* for the middle range of n, and hence resulted in tighter bounds than those given by Elias' end formulas. These numerical results are presented in the section on "Bounds on the Probability of Error" and, for specific cases, are useful principally as guides in interpreting Elias' bounds.

The most interesting results are presented in the Section on the "Evaluation of Actual Codes on the 704." The IBM *704* computer was programmed to evaluate probability of error for binary group alphabets with n as large as 72 (and with no more than 20 parity check bits). The *704* was used for two general purposes: first, to aid in proving more codes to be optimum, thus extending Slepian's table; and secondly, to aid in studying random codes.

The possibility of using random codes has often been mentioned. Most randon codes are good since the average error probability for random codes is in general near optimum. The real difficulty lies not in generating a random code, but rather in determining just how good a specific random code is. Now, using the *704*, it is possible to generate fairly long random codes, evaluate them, and select codes which are good.

Review of Binary Coding Theory

The pertinent aspects of the theory of binary codes are reviewed briefly here. The codes with which this paper is concerned are block codes for the binary symmetric channel. The binary symmetric channel is a communication channel which transmits two symbols, represented by the binary digits *0* and *1*, with equal probability q that the received symbol will be the same as the transmitted symbol, and probability p that the received symbol will be the opposite of what was transmitted, i.e., that an error will occur. The errors are assumed to be independent.

A block code is a code using sequences of binary digits, *1*'s and *0*'s, of a fixed length n. Only certain selected sequences, called code points, are transmitted. During transmission certain of these binary digits may change, i.e., errors may occur. If a given code word is transmitted, the probability that it will be received with no error is q^n. The probability that there will be one error in a specified position is pq^{n-1}. The probability of i errors in i specified positions is p^iq^{n-i}. If $q>p$, the received block with no errors is more likely than any specified sequence with one error. Any received sequence with one error is more likely than any specified received sequence with two errors, etc.

At the receiver a decision must be made as to which code point was transmitted, i.e., the received sequence must be decoded. This might be done according to the system illustrated in Fig. 1. The code points are listed in a row across the top of the table, and under each code point are listed all other sequences of n binary digits which will be "decoded" into the given code point.

The probability of error for a code can be calculated from a table such as that shown in Fig. 1. The probability of correct decoding if a given code point is transmitted equals the probability that the received sequence is one listed under the transmitted code word. For example, consider column two in Fig. 1. The code point itself involves no errors, and hence has probability q^n. The next five sequences down the list result from single errors and hence have probability pq^{n-1}. The last two are double errors and hence have probability p^2q^{n-2}. Thus the total probability of correct decoding if this code point is transmitted is

$$Q=q^n+5pq^{n-1}+2p^2q^{n-2}$$

For the code as a whole the probability of error is the average probability of error for all the code points. If the code points are used with equal frequency, this is the sum of the error probabilities for the individual code points, divided by the number of code points.

Block codes can be subdivided into parity check and nonparity check codes. For the most part this paper deals with parity check codes. In such a code k positions are used as information positions and in them all combinations of ones and zeros are permitted. The symbols in the remaining $n-k$ positions are determined from the information bits by "parity checks." The code shown in Fig. 1 is a parity check code. The first two places are information bits. The fourth bit is a parity check on both information places—it is one if there are an odd number of ones in information places, zero if there are an even number of ones in information places. The third bit is a parity check on the first bit only, the fifth bit a parity check on the second bit only.

For a parity check code the optimum decoding scheme is as follows: The decoding table (for example, Fig. 1) when completed will list each sequence once and only once. The code points are listed in the first row with the word consisting of all zeros first. Then of all the remaining sequences the one with the fewest ones in it (or any one of these if there are several) is placed under the code point consisting of all zeros. This sequence represents an error pattern, since each *1* in the sequence represents an error in that digit. Next under each of the other code points is listed the sequence which would result if the same error pattern should occur with the code point in question. The row of sequences formed in this manner is called a "coset," and the sequence under the first code point (00000) is the "coset leader."

A third row is formed using as its coset leader an unused sequence with the fewest ones, and similarly a fourth row, fifth row, and so on until all sequences of n binary digits are used. The complete table is called a "standard array." In the basic theory of "group alphabets" described by Slepian, it is proved that such a scheme is without duplications and represents optimum decoding.

Since the sets of error patterns which are corrected are the same for all code points in a parity check alphabet, the probability of correct decoding is the same for all code points. In fact, this probability can be calculated if only the sequences in the first column, i.e., the coset leaders, are known. For example, in the code shown in Fig. 1, the first sequence in column one has weight zero, i.e., it has no 1's. The next five sequences have weight one, and the last two have weight two. The probability of correct decoding for the code point of all zeros, and therefore for every code point, is

$$Q = q^5 + 5pq^4 + 2p^2q^3$$

Since a parity check code has k information places each of which may be zero or one, there must be 2^k code points. There are 2^n different sequences of n binary digits. Hence there must be 2^{n-k} cosets.

Bounds on the Probability of Error

The probability of error P_{opt} for an optimum parity check code for the binary symmetric channel, given the probability of error for the channel, the length of the transmitted sequences, and the number of information digits in a transmitted sequence can be bounded by $P_t \leq P_{opt} \leq P_{av}$. Recent papers of Elias[3-6] give expressions for P_t and P_{av}. To simplify these expressions, Elias obtains upper and lower bounds and asymptotic expressions for P_t and P_{av}. From the bounds for P_t and for P_{av}, he draws some very interesting and important conclusions regarding the exponential behavior of P_{opt} for large n. The meaning of the bounds P_t and P_{av} will be reviewed in this section. In addition, values calculated from a lower bound for P_t will be compared with those calculated from the exact expression for P_t; and values computed from the upper bound for P_{av} will be compared with those computed from a lower upper bound for P_{av} to give a clearer idea of the range of usefulness of these expressions. There is also an upper bound for P_{opt} which is due to Gilbert.[10] This also will be discussed and evaluated for several values of n, k, and p.

THE LOWER BOUND P_t

Suppose that for a given code and decoding scheme the list of sequences which decode into each code word includes all error patterns with fewer than j errors, some error patterns with j errors, but none with more than j errors. Then moving one sequence from the column under one code word to the column under another decreases the probability of correct decoding by p^jq^{n-j} for the first word, and increases the probability of correct decoding by no more than this amount for the second code word. This procedure cannot possibly increase the average probability of error (if the code words are used with equal frequency). Thus such a code is clearly as good a code as can be found with the same number of code words and block length. The code described in Fig. 1 is such a code.

Such codes do not exist for most values of block length and number of code points. Whether they do or not, clearly no code could have a greater probability of correct decoding than would be calculated, assuming such a code were possible. This fact provides the lower bound on the probability of error found by Elias:

$$P_t = 1 - \sum_{j=0}^{k_1-1} \binom{n}{j} p^j q^{n-j} + \left[2^{n-k} - \sum_{j=0}^{k_1-1} \binom{n}{j}\right] p^{k_1} q^{n-k_1} \tag{1}$$

where k_1 is the integer defined by

$$\sum_{j=0}^{k_1} \binom{n}{j} \geq 2^{n-k} > \sum_{j=0}^{k_1-1} \binom{n}{j} \tag{2}$$

Elias' bounds for P_t are

$$P_t \geq \frac{q_1 p}{\left(p_1 + \frac{1}{n}\right) 2q\sqrt{2np_1q_1}} \times 2^{-n[H(p^1) - H(p) + (p_1 - p) \log q/p]} \tag{3}$$

$$P_t \geq \frac{p_1 q}{(p_1 - p)\sqrt{2\pi n p_1 q_1}} \times 2^{-n[H(p_1) - H(p_1) + (p_1 - p) \log q/p]} \tag{4}$$

where $p_1 = k_1/n$ and $p_1 + q_1 = 1$ and $H(p) = -p \log p - q \log q$

THE UPPER BOUND P_{av}

A random code is formed by selecting the code points at random from the possible 2^n sequences. The code points are selected independently with all of the 2^n sequences equally probable. Received sequences are decoded into a code point which differs in the fewest positions from the received sequence. The average probability of error for random codes is defined as P_{av}. It is found by averaging the probability of error for the decoded word over all code points in any particular code and over all possible codes, with equal weights given to each code point and to each code. P_{av} provides an upper bound for P_{opt} since at least one of the random codes must be as good as the average. If $P_{av}(j)$ is defined as the average probability of error for random codes when j errors are made in transmission, the following expression can be written for P_{av}:

$$P_{av} = \sum_{j=0}^{n} P_{av}(j) \binom{n}{j} p^j q^{n-j} \tag{5}$$

since

$$\binom{n}{j} q^j q^{n-j}$$

is the probability of j errors. To simplify equation 5, Elias obtains the following upper bounds for $P_{av}(j)$

$$P_{av}(j) \leq \frac{\sum_{i=0}^{j} \binom{n}{i}}{2^{n-k}} \leq \frac{\binom{n}{j}}{\binom{n}{k_1-1}} \text{ for } j \leq k_1 - 1$$

$$P_{av}(j) \leq 1 \text{ for } j \geq k_1$$

The expression for P_{av} now becomes

$$P_{av} < P_{av}' = \sum_{j=k_1}^{n} \binom{n}{j} p^j q^{n-j} + \sum_{i=0}^{k_1-1} \frac{\sum_{i=0}^{j} \binom{n}{j}}{2^{n-k}} \binom{n}{j} p^j q^{n-j} \tag{6}$$

$$P_{av}' \leq P_t + \sum_{j=0}^{k_1-1} \frac{\binom{n}{j}^2}{\binom{n}{k_1-1}} p^j q^{n-j} \tag{7}$$

The symbol P_{av}' is introduced only to have an easy reference to equation 6. Finally, Elias defines p_c and q_c which are functions of p and q:

$$p_c = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{q}} \qquad q_c = \frac{\sqrt{q}}{\sqrt{p} + \sqrt{q}}$$

and obtains the following two bounds for P_{av}': if $p_1 < p_c$

$$P_{av}' < P_t^{ub} + \frac{p_1 q}{q_1 p \left[1 - \left(\frac{p_1}{q_1}\right)^2 \frac{q}{p}\right] \sqrt{2\pi n p_1 q_1}} \times 2^{-n[H(p) - H(p_1) + (p_1 - p) \log q/p]} \tag{8}$$

and if $p_1 > p_c$,

$$P_{av}' \leq p_t^{ub} + \frac{\left(q_1 + \frac{1}{n}\right)\sqrt{2np_1q_1}}{\pi p_c q_c} \times 2^{-n[H(p) + H(p_1) - 2H(p_c) + (p_c - p) \log q/p]} \tag{9}$$

where P_t^{ub} is the upper bound (equation 4) for P_t. The preceding expressions (equations 6, 8, and 9) for P_{av}' are for average of all codes. Elias[3,4] proves that the same expressions for P_{av}' are also valid for average of all parity check codes. However, the P_{av} for all codes may differ from the P_{av} for parity check codes, since P_{av}' is only an upper bound for P_{av}.

THE UPPER BOUND P_g

Gilbert[10] has discussed a lower bound for the maximum number of n-symbol binary sequences for which every pair differs in at least d positions. His argument can be rephrased to show that there is a set of 2^k n-digit sequences every pair of which differs in at least d positions, where d is the integer satisfying

Vol.
Pt. 1
195

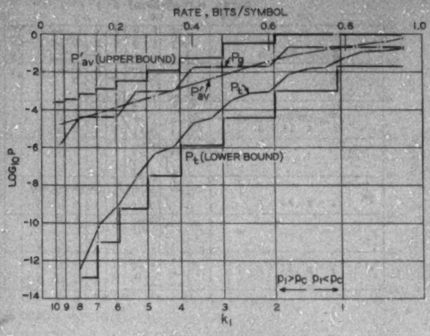


Fig. 2. Bounds for n=20, p=0.01

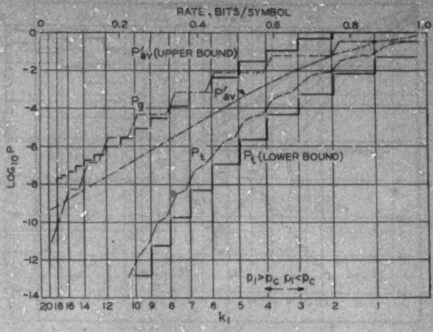


Fig. 4. Bounds for n=40, p=0.01

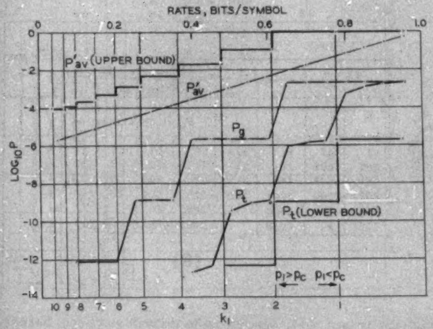


Fig. 3. Bounds for n=20, p=0.0001

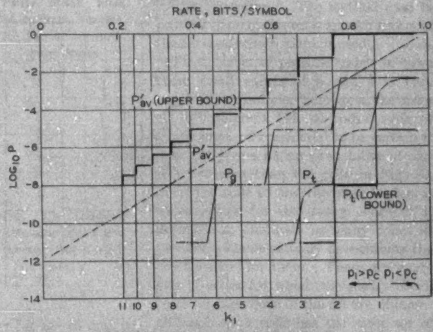


Fig. 5. Bounds for n=40, p=0.0001

$$2^k\sum_{j=1}^{d}\binom{n}{j} > 2^n \geq 2^k\sum_{j=0}^{d-1}\binom{n}{j}$$

If every pair of the 2^k code points differs in at least d positions, it is possible to correct all error patterns with fewer than $d/2$ errors. Thus, Gilbert's argument provides an upper bound for the probability of error which will be called P_g:

$$P_g = 1-\sum_{j=0}^{k_2}\binom{n}{j}p^jq^{n-j} \tag{10}$$

where k_2 is the integer satisfying $k_2 < d/2 \leq k_2+1$. Elias in his unpublished paper, "Coding for Two Noisy Channels," has noted that Gilbert's bound applies to parity check codes as well as nonparity check block codes so that $P_{opt} \leq P_g$.

CALCULATED RESULTS

In this section the exact expression for P_t (equation 1) will be compared with Elias' lower bound for P_t (equation 3). Also the exact expression for P_{av}' (equation 6) will be compared with Elias' upper bound for P_{av}' (equations 8 and 9) and with P_g (equation 10).

In Figs. 2–9, the logarithm of the probability of error for the various bounds is plotted as a function of rate for values of n equal to 20, 40, 80, and 160 and values of p equal to 0.01 and 0.0001.

The equations for P_t, P_{av}', and P_g are evaluated for integral values of k. Strictly speaking, this gives a series of points at the values of rate equal to k/n; but to indicate which points belong to which bound, the points are connected by straight lines. For n equal to 20 and 40, all the points are shown. For n equal to 80 and 160, the only points shown are the ones for which there is an abrupt change in the value of the equation. The omitted points lie either on or very close to the straight line segments.

Since k_1 and not k enters directly in equations 3, 8, and 9, it is convenient to evaluate them using integral values of k_1.

The rate R can then be calculated from equation 2 as:

$$R = k/n = 1-\frac{1}{n}\log\sum_{j=0}^{k_1}\binom{n}{j} \tag{11}$$

In most instances this does not give a value for the rate for which nR and hence k is an integer, but an evaluation of equation 3 at this value of rate provides a lower bound for P_{opt} for all rates larger than this. Similarly, evaluating equations 8 or 9 for a given k_1 and computing the corresponding rate from equation 11 gives an upper bound for P_{opt} for all rates less than this amount. The values of k_1 are shown in the figures. Also indicated are the values of k_1 for which $p_1 = k_1/n < p_c$ and for which $p_1 > p_c$.

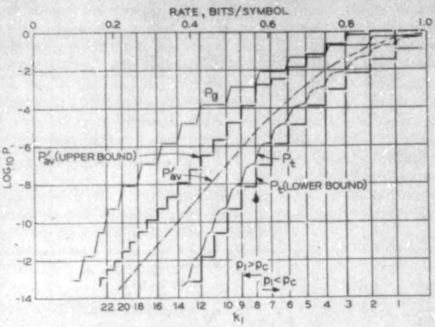


Fig. 6. Bounds for n=80, p=0.01

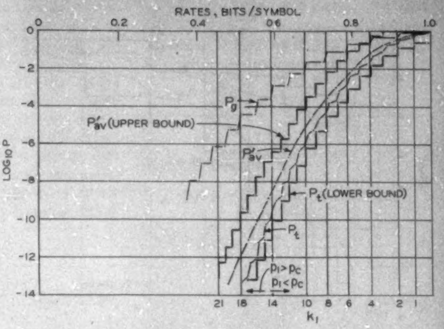


Fig. 8. Bounds for n=160, p=0.01

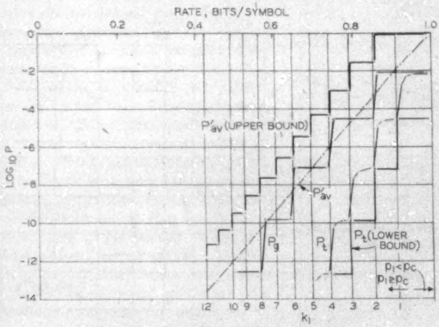


Fig. 7. Bounds for n=80, p=0.0001

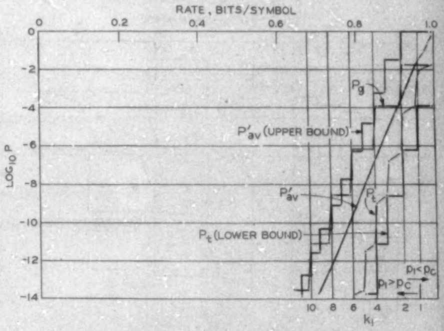


Fig. 9. Bounds for n=160, p=0.0001

DISCUSSION

Several observations can be made from the figures. For the most part, P_{av}' and P_t differ by several orders of magnitude for these small block lengths. The calculated values of P_{av}' and P_{av}' (upper bound) differ by a factor of ten or more, with P_{av}' the smaller. The calculated values for P_t are one or two orders of magnitude larger than P_t (lower bound) and this difference does not depend strongly on the rate but becomes smaller as n increases. For $p=0.0001$, P_g is a better upper bound than P_{av} for most values of rate except when $n=160$.

It is interesting that many of Elias' conclusions regarding the bounds for large n are applicable for these small values of n, and that when the curves are interpreted using the k_1 scale, Elias' conclusions are still substantially correct for these small values of n. For example, one of Elias' observations was that P_{av}' and P_t have the same exponent for $p_1<p_c$ and their ratio approaches a constant for large n. The curves for $p=0.01$ in the region for which $p_1<p_c$ exhibit this same property. Another conclusion was that as $p_1 \rightarrow p$, which is the same as the rate approaching the channel capacity, the ratio of the asymptotic expressions for P_{av}' and P_t approaches unity. The ratio of P_{av}' to P_t is nearly unity as $p_1 \rightarrow p$ for the curves that are shown, provided np_c is larger than one or two. If np_c is less than one, P_{av}' remains one or two orders of magnitude larger than P_t as $p_1 \rightarrow p$. For large n, random coding is as good as any coding scheme for rates approaching channel capacity. The figures show that this is not necessarily true for the bounds when n is small although it may be true for the actual codes. Finally, Elias concludes that for any $p_1<p_c$, random coding is very nearly as good as anything else, while for very low rates and sufficiently large n, P_g will be a lower upper bound than P_{av}'. These statements still hold for small n.

Evaluation of Actual Codes on the 704

DISCUSSION OF THE 704 PROGRAM

The operation of the program is based on two properties of the parity check sequence. The parity check sequence for a given received sequence is an $n-k$ symbol sequence of zeros and ones obtained as follows: The parity checks for the received sequence are calculated by using the information digits in that sequence. If the parity check computed from the received information symbols agrees with the corresponding parity check digit in the received sequence, a zero is written for the corresponding digit in the parity check sequence. If the two parity check symbols disagree, a one is written. Slepian

$$\begin{array}{c|cccccccc} & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 \\ x_1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ x_2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ x_3 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ x_4 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{array} \qquad n = 12,\ k = 4$$

$$\begin{array}{ccccccccc} 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \end{array} \qquad n = 13,\ k = 4$$

$$\begin{array}{cccccccccc} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \end{array} \qquad n = 14,\ k = 4$$

$$\begin{array}{ccccccccccc} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{array} \qquad n = 15,\ k = 4$$

Example: $n = 12$, $k = 4$

Message	Transmitted Sequence
$x_1x_2x_3x_4$	$x_1x_2x_3x_4y_1y_2y_3y_4y_5y_6y_7y_8$
1 1 0 0	1 1 0 0 0 1 1 1 1 0 0 0

(A)

i	n = 12		n = 13		n = 14		n = 15	
	α_i	$\binom{12}{i}$	α_i	$\binom{13}{i}$	α_i	$\binom{14}{i}$	α_i	$\binom{15}{i}$
0	1	1	1	1	1	1	1	1
1	12	12	13	13	14	14	15	15
2	66	66	78	78	91	91	105	105
3	144	220	246	286	364	364	455	455
4	33	495	173	715	511	1001	993	1365
5			1	1287	42	2002	475	3003
6					1	3003	4	5005

(B)

Fig. 10 (left). (A) Parity check rules of optimum codes. (B) The α's for optimum codes

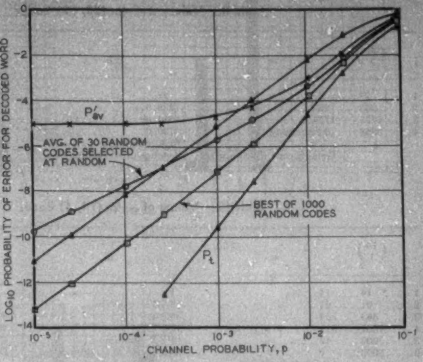


Fig. 11 (above). Random codes and bounds for $n=36$, $k=19$

showed that every member of a coset has the same parity check sequence and that no two cosets have the same parity check sequence.[2]

Given the parity check rules for the code, the program finds the 2^{n-k} coset leaders by trial and error. All sequences of weight one are tried first as proposed coset leaders. Then all sequences of weight two are considered and so forth. The parity check sequence is computed for each new proposed leader. If the parity check sequence has been used previously, the proposed leader cannot be used since it is in a coset whose leader was found previously. Otherwise, the proposed leader can be listed as a coset leader and its parity check sequence recorded. Since the weights of the successive proposed coset leaders increase monotonically, no element of a coset has smaller weight than the leader of that coset.

There are actually two programs. The first requires that $n \leq 36$ and $n-k \leq 20$ and the second that $37 \leq n \leq 72$ and $n-k \leq 20$. The first tests approximately 16,000 possible leaders per minute for $n=36$ and is somewhat faster for smaller n. The coset leaders with weight four or less can be found for a $n=14$, $k=4$ (14, 4) code in four seconds and for a $n=36$, $k=19$ (36, 19) code in four minutes. To get an adequate estimate of the probability of correct decoding for codes within the range of the program, it is usually not necessary to find the number of coset leaders with weight greater than four or five. The evaluation can be stopped after all sequences of an arbitrary weight are tried, thereby saving considerable machine time.

Application to Finding New Optimum Group Codes

Using this program we have extended Slepian's table by proving that the codes whose parity check rules are shown in Fig. 10 are optimum codes for $k=4$ and $n=12$, 13, 14, and 15. In Fig. 10(B), α_i is the number of error patterns with exactly i errors in them that a code will correct. Thus α_i is the number of coset leaders containing exactly i ones. These codes were constructed according to MacDonald's coding procedure.[9] The method used in proving them optimum is an adaptation of the method which Slepian used.

The number of possible codes for even these small values of n and k is very large. For example, there are in the order of 10^{12} different (14, 4) codes, each of which must be considered to be certain it does not give a higher probability of correct decoding than the code in Fig. 10. Since the time to find the coset leaders is unreasonable for this many codes, it is desirable to eliminate as many codes as possible without actually determining the coset leaders for them. There are many sets of codes for which the parity check rules for each code of the set are different although each code has the same set of α_i's. The *704* program was written to test at least one code while ignoring many, but not necessarily all, of the remaining codes in every such set. Secondly, the weights of the code points for a given code can be computed in a fraction of the time required to evaluate the code. It is possible to find a lower bound for the weight of all code points of any code that is better than the assumed optimum. Thus by computing the minimum weight of the code points and comparing it with the lower bound for the minimum weight, it was possible to eliminate many codes

Table I. Values of α_i for (36, 19) Codes

i	$\binom{36}{i}$	α_i	α_i
1	36	36	36
2	630	628.4	630
3	7140	6923.6	7080
4	58905	44914.2	47481
5	376992	(78568.8 remaining)	73156 (2688 remaining)
		Average of 30 random codes	Best of 1,000 Random codes

Table II. Values of α_i for (23, 12) Codes

i	$\binom{23}{i}$	α_i	α_i	α_i
1	23	23	23	23
2	253	253	238.6	253
3	1771	1771	1046.0	1199
4	8855	(0 remaining)	730.0	572
			(9.0 remaining)	(0 remaining)
		Golay's code (optimum)	Average of 25 random codes	Best of 3,000 random codes

Table III. Values of α_i for (14, 4) Codes

	$\binom{14}{i}$	α_i	α_i	α_i
1	14	14	14	14
2	91	91	87.3	91
3	364	364	285.4	344
4	1001	511	422.7	496
5	2002	42	(213.7 remaining)	(78 remaining)
6	3003	1	Average of 25 random codes	Best of 2,000 random codes
		(0 remaining) optimum		

without actually finding their α_i's. For the (14, 4) code this left 5,465 codes for which the α_i's had to be found. Approximately $5^1/_2$ hours of *704* machine time was needed to prove the (14, 4) code, shown in Fig. 10, to be optimum.

APPLICATION TO RANDOM CODES

Another application for the code evaluation program is random group codes. Several questions of interest concerning random codes are: 1. How conservative is the upper bound P_{av}' obtained in the section on "Bounds on the Probability of Error," compared to the actual average performance of random codes? 2. How good a code can be obtained by taking the best of several codes selected at random? and 3. How are random codes distributed about the average? In this section we shall present some experimental data pertinent to these questions.

The random codes were generated by using pseudo-random numbers to determine parity check rules. The actual procedure[11] was to use the $n-k$ high-order digits of the binary representation of successive powers of 5^{13} mod 2^{35}. One such set of digits is used for each of the k information digits: a one in the ith position in the $n-k$ random digits is taken to mean that the ith parity check includes the information digit in question.

Thirty random (36, 19) codes were generated and each evaluated to find the α_i's from α_0 through α_4. This left some coset leaders undetermined. The probability of error for the decoded word is

$$P=1-\sum_{i=0}^{n}\alpha_i p^i q^{n-i}$$

An upper bound on this probability can be found by neglecting coset leaders with weight greater than four. A lower bound is obtained by considering that all coset leaders with weight greater than four have weight five. Since for values of p equal to 0.1 or less these two bounds would be indistinguishable in Fig. 11 if both were plotted, the coset leaders with weight greater than four were not determined. The polynomial for the probability of error is linear in the α_i's. Therefore, the average probability of error for the 30 codes can be obtained by using the average values of the α_i's for the 30 codes. These average values are tabulated in Table I. In Fig. 11, the probability of error for the decoded word for the average of the 30 random codes is plotted against the probability of error for the channel. Also plotted are the P_g, P_t, and P_{av}' bounds. The probability of error for the decoded word for the average of these 30 codes is always less than P_{av}', which is to be expected, since P_{av}' is an upper bound for the average probability of error for random codes. P_{av} is a good upper bound for the probability of error for random codes at large values of p but a poor upper bound for small values of p. This would indicate that the average probability of error for random codes for the conditions of Fig. 5, for example, is much closer to P_t than P_{av} indicates. It is known from theory that P_g is a better bound than P_{av} for low rates. It is interesting to note that at low values of p, P_g is also a better bound than the average of the 30 codes. This might be expected, since for very low channel probability of error, the best code has a maximum minimum distance. Gilbert's bound is based on an estimate of minimum distance, but random codes would not in general be expected to have a large minimum distance.

To see how good a code can be obtained by generating random codes and selecting

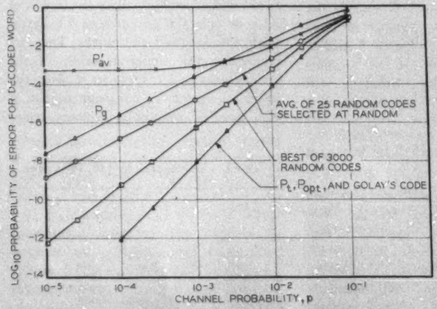


Fig. 12. Random, optimum, and Golay's codes and bounds for n = 23, k = 12

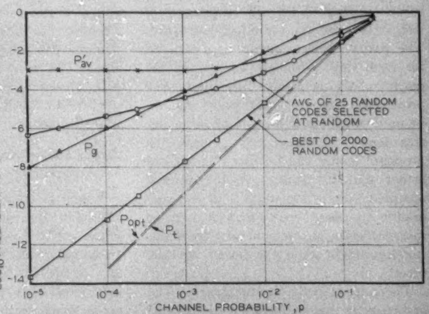


Fig. 13. Random and optimum codes and bounds for n = 14, k = 4

$$\begin{array}{cccccccccccc}
0&0&1&1&1&1&1&0&0&1&0&0\\
0&1&1&0&0&0&1&1&1&1&1&1\\
1&0&0&1&1&0&0&0&0&0&1&0\\
1&1&1&1&0&0&1&1&1&0&1&0\\
0&1&0&1&0&1&1&1&1&1&0&1\\
1&1&0&1&1&1&0&0&1&1&0&0\\
1&1&1&1&1&0&1&1&0&0&0&1\\
1&0&0&1&1&0&1&1&0&0&1&1\\
0&0&1&1&0&0&1&0&1&1&0&0\\
1&1&0&1&0&1&1&1&0&0&0&1\\
1&1&0&1&1&0&1&0&1&0&1&1\\
0&1&1&0&0&1&1&1&1&0&0&1\\
0&1&1&1&1&0&1&0&1&0&0&1\\
1&0&0&1&1&1&0&0&1&1&1&1\\
1&1&1&1&0&1&1&0&1&0&0&1\\
0&0&1&1&0&1&1&1&0&0&1&1
\end{array}$$

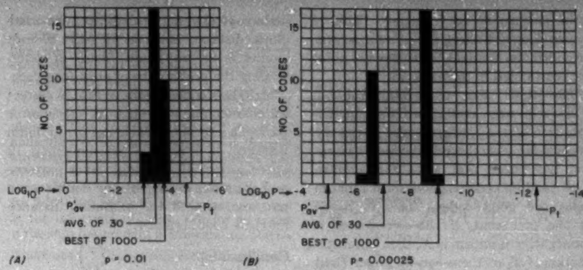
(A)

i	α_i	$\binom{28}{i}$
1	28	28
2	378	378
3	2287 (1402 Remaining)	3276

(B)

Fig. 14. A (28, 16) code that corrects all single, double, and adjacent triple error patterns'

A—Parity check rules
B—Weight and number of corrected error patterns



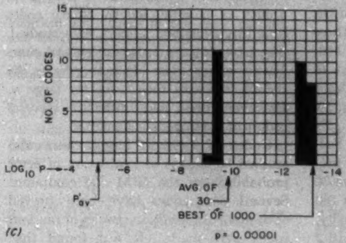


Fig. 15. Distribution of probability of error after decoding for 30 random (36, 19) codes. Channel probability equals 0.01, 0.00025, 0.00001 (one square per code)

the best, 1,000 random (36, 19) codes were generated and evaluated. The code that would correct all the single and double errors and the most triple errors was considered best. This gave a code considerably better than the average as shown by the α_i's tabulated in Table I and the probability of error for the decoded word plotted in Fig. 11. The curve for this best random code lies midway between the curves for the average of 30 codes and P_t and provides a good upper bound for the optimum (36, 19) parity check code. The optimum (36, 19) parity check code is not known and probably does not achieve the bound P_t, so that the best of the 1,000 random codes may be even closer to P_{opt} than is indicated by its nearness to P_t.

Except for Hamming's codes, Golay's "lossless" (23, 12) code is the largest known optimum code. Since this code corrects all single, double, and triple errors and no others, it achieves the bound P_t. To see how good the best of several random codes was in comparison with Golay's code, 3,000 (23, 12) random codes were generated and evaluated. The α_i's for the best code of these 3,000 are tabulated in Table II, and the probability of error for the decoded word is plotted in Fig. 12. Although the best of 3,000 random codes is much better than the bounds P_{av}' and P_g, it is considerably poorer than Golay's code.

Most optimum codes are not lossless codes, and optimum codes for many combinations of n and k do not achieve the lower bound P_t. To see how random coding compares with an optimum code which does not achieve P_t, several (14, 4) random codes were generated and evaluated. The α_i's for the average of 25 (14, 4) random codes selected at random and for the best of 2,000 (14, 4) random codes are tabulated in Table III. The code that corrected all the single- and double-error patterns and the most triples was considered best. The probability of error for the average of the 25 codes, for the best of the 2,000 codes, and for the optimum (14, 4) code together with the bounds are shown as functions of the channel probability in Fig. 13. The optimum (14, 4) code differs from the nonexistant hypothetical code represented by P_t in that the optimum code corrects 43 fewer quadruple-error patterns; instead it corrects 42 quintuple and one sextuple patterns. This has very little effect on the probability of error for the decoded word as is shown in Fig. 13. The best of 2,000 random codes is considerably better than the average of the 25 random codes and the bound P_g. It is interesting that for a given value of channel probability the best (14, 4) and (23, 12) random codes that were found differ from the optimum code by approximately the same amount.

In addition to being used to find a good code for a given n and k, random codes can also be used to find codes which will correct certain error patterns. For example, suppose that it is desirable to have a code that corrects all error patterns with one and two errors in them and at least all error patterns with triple errors that are in adjacent digit positions. The parity checks and the α_i's for a (28, 16) code that does this are shown in Fig. 14. The code was found by generating several random codes and saving only those that corrected all single- and double-error patterns. These were then evaluated to find out how many triple-error patterns were corrected. The evaluation program was modified to try next the 26 adjacent triple-error patterns as coset leaders. If the code were unable to correct one of these, the program tried to find a rearrangement of the symbols in the code points that would allow the triple-error pattern to be corrected. This of course does not affect the α_i's for a code, provided the symbols for each code point are rearranged in the same way.

The third area concerning random codes that was investigated was the distribution of the random codes about the average for several random codes. Shannon has noted that since the average of all random codes is good, most random codes are good codes. This is based on the fact that, if the average of n nonnegative numbers is X, then at most a fraction, $1/f$, of them can be larger than fX. The probability of error for the decoded word was calculated for each of the 30 (36, 19) random codes for three values of channel probability (0.01, 0.00025, 0.00001). The distribution of the probability of error for the 30 codes is shown as a frequency bargraph in Fig. 15. Also shown are the bounds P_{av}, P_g, P_t, and the probability of error for the best of 1,000 random codes. From the preceding, there could be at most 3 of the 30 codes that have a probability of error ten times larger than the average. Actually, there are none. However, for $p=10^{-5}$ there are 18 of the codes that have a probability of error two or three orders of magnitude smaller than the average, and of these the best is nearly equal to the best of 1,000. Thus, the best of 20 or 30 random codes selected at random is a good code and it appears that the improvement gained in using a larger number than this is small unless, perhaps the number is very large.

Fig. 15 illustrates again that for small values of channel probability, the minimum distance between all pairs of code points is an important factor in determining the probability of error for a code. The 18 best codes shown in Fig. 15 correct all 630 double but not all triple-error patterns and hence each has a minimum distance of 5 or 6. The remaining 12 codes correct all the single-error patterns and from 621 to 627 of the double-error patterns, so that the minimum distance for these codes is 3 or 4.

Conclusions

Upper and lower bounds on error probability of medium length binary codes found by Elias and Gilbert were studied. The bounds have the same general characteristics for these block lengths as the theory indicates they have for large n. However, they may differ by several orders of magnitude for small n.

Codes of medium length were also studied by means of evaluation of error probability on the IBM *704* computer. Several new codes have been proved optimum. In addition studies of random codes were made. It was found that the upper bound for the average probability of error for all random codes may be three or four orders of magnitude larger than the actual average for random codes for small p. It was also observed that good codes can be obtained by generating and evaluating several random codes and selecting the best of these. In the cases investigated, there was little advantage in generating more than 20 or 30 codes to obtain a good one. Although it is easy to obtain a good code using random codes, the best random codes in 3,000 (23, 12) and 2,000 (14, 4) codes were not as good as the optimum for these values. Finally, it was possible to use random codes to obtain codes which correct certain specific error patterns.

References

1. The Mathematical Theory of Communication, C. E. Shannon. University of Illinois Press, Urbana, Ill., 1949.

2. A Class of Binary Signalling Alphabets, D. Slepian. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1956, pp. 203–34.

3. Coding for Noisy Channels, P. Elias. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1955, pp. 37–46.

4. Coding for Two Noisy Channels, P. Elias. "Information Theory," Colin Cherry, editor. Butterworth's Scientific Publications, London, England, 1956, pp. 61–74.

5. List Decoding for Noisy Channels, P. Elias. *WESCON Convention Record*, Institute of Radio Engineers, pt. 2, 1957, pp. 94–104.

6. Notes on Digital Coding, M. J. E. Golay. *Proceedings*, Institute of Radio Engineers, vol. 37, 1949, p. 657.

7. Error Detecting and Error Correcting Codes, R. W. Hamming. *Bell System Technical Journal*, vol. 29, 1950, pp. 147–60.

8. A Class of Multiple Error-Correcting Codes and the Decoding Scheme, I. S. Reed. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, 1954, pp. 38–49.

9. Constructive Coding Methods for the Binary Symmetric Independent Data Transmission Channel, J. E. MacDonald. *MEE Thesis*, Syracuse University, Syracuse, N. Y., 1958.

10. A Comparison of Signalling Alphabets, E. N. Gilbert. *Bell System Technical Journal*, vol. 31, 1952, pp. 504–22.

11. Generation and Testing of Pseudo-Random Numbers, O. Taussky, J. Todd. "Symposium on Monte Carlo Methods," H. A. Meyer, editor. John Wiley & Sons, Inc., New York, N. Y., 1956, pp. 15–28.

◆

Discussion

The following closure includes a number of oral comments by Professor P. Elias on the paper and the authors' reply to them.

A. B. Fontaine and W. W. Peterson: Prof. Elias pointed out to us that the expression P_{av}' is greater than P_{av}, the average probability of error or ambiguity for all block codes, by no more than about 60%. An exact expression for $P_{av}(j)$ is:

$$P_{av}(j) = 1 - \left[1 - 2^{-n}\sum_{i=0}^{j}\binom{n}{i}\right]^{2^k-1} \tag{12}$$

The "well-known inequality" $1-(1-x)^y \leqq xy$ is then applied to give

$$P_{av}(j) \leqq \frac{2^k-1}{2^n}\sum_{i=0}^{j}\binom{n}{i} \tag{13}$$

or 1, whichever is smaller. The greatest difference in terms of percentage occurs when the right-hand side of equation 13 is unity, and that can easily be shown to be less than $e/(e-1)=1.582$. Dropping the one in the factor 2^k-1 increases the right-hand side by one part in 2^k, which is usually negligible.

Elias has shown that P_{av}' is an upper bound on the average probability of error for all parity-check codes. However, more approximations are involved, so that P_{av}' is not a close upper bound for small channel probability of error p. In fact, the worst possible parity-check code is the code in which the parity checks check no information places, and it has the same probability of error as if there were no parity checks at all, i.e., $1-(1-p)^k$. This approaches zero (approximately as kp) as p approaches zero while P_{av}' approaches $2^{-(n-k)}$. A closer bound on the average probability of error for all parity check codes, derived below, is

$$P_{av}'' = 2^{-(n-k)}\sum_{j=1}^{n} p^j q^{n-j}\left[\binom{n}{j}\sum_{i=0}^{j}\binom{n}{i} - \sum_{r=0}^{j}\sum_{i=0}^{j}\binom{n-k}{j-r}\binom{n-k}{i-r}\binom{k}{r}\right] \tag{14}$$

As p approaches zero, all terms beyond the first in equation 14 become negligible, and

$$P_{av}'' \approx 2^{-(n-k)}kp(2n-k) \tag{15}$$

It can be shown that the correct value of the leading term in the average probability of error or ambiguity for all parity-check codes is

$$2^{-(n-k)}kp[2n-k-2^{-(n-k)}(k-1)(n-k+1)] \tag{16}$$

which differs very little from equation 15 unless $n-k$ is very small.

When compared to the observed average for the samples of 20 or 30 random (23, 12) and (36, 19) codes, the discrepancies are much greater than this. Elias suggested the following explanation: There are not many very bad codes, but they may contribute a great deal to the average probability of error, because their probability of error (for small p) may be several orders of magnitude greater than the average. A small sample is not likely to include one of these, and therefore, has a high probability of having a sample average smaller than actual average. In short, the distribution of probability of error is so skew that the distribution of sample average is also very skew.

The upper bound on average probability of error for parity-check codes will be derived assuming the reader familiar with the theory of group codes.[1] A parity-check code can be described by giving its generators in the form of an $n \times k$ matrix of ones and zeros. The first k columns at the left are called the information places, and in the $n \times k$ matrix they form a $k \times k$ identity matrix. The remaining $n-k$ columns are parity-check columns, and the configuration of ones and zeros determines the parity-check rules. Each of the $2^{k(n-k)}$ ways of filling ones and zeros in these $k(n-k)$ positions describes a different parity-check code, though many differ in trivial ways. An upper bound is desired for the average probability of error for all these codes, weighting them equally.

As a first step, the number of codes containing a given n-bit sequence as a code word is required. If the n-bit sequence has all zeros in the information places, but does not have all zeros in the parity-check places, it cannot be a code word in any code. If, however, it does not have all zeros in the information places, then there are $2^{(k-1)(n-k)}$ codes containing this word. This can be seen as follows: if the given sequence has more than a single one in the information places, generators should be added (bit by bit, modulo 2) until there is only a single bit left in the information places. The n-bit sequence derived in this manner is a code word if, and only if, the original sequence was. Since it has only a single one in the information places, it must be a generator. The other $k-1$ generators are arbitrary, but once they are chosen this last generator is also determined. Thus, $(k-1)(n-k)$ bits can be chosen freely, and $2^{(k-1)(n-k)}$ codes constructed which contain the given word as a code word.

Now consider an error pattern which has j ones of which r are in the information places. This error pattern will result in ambiguity or erroneous decoding if there is any sequence containing $i \leq j$ ones such that the sum of the error pattern and the other sequence is a code point. This will happen for $2^{(k-1)(n-k)}$ codes if the sum has any ones in the information places, but for no codes if the sum has no ones in information places. There are $\binom{n}{i}$ sequences containing i ones, of which $\binom{n-k}{i-r}$ have information places identical with those in the error pattern, and hence when added to the error pattern results in a sequence with all zeros in the information places. This leaves $\binom{n}{i}-\binom{n-k}{i-r}$ sequences with i ones for which the sum does not have all zeros in the information places. Therefore, there are at most

$$2^{(k-1)(n-k)}\sum_{i=1}^{j}\left[\binom{n}{i}-\binom{n-k}{i-r}\right] \qquad (17)$$

codes out of the total of $2^{k(n-k)}$ which fail to correct a particular error pattern with j errors of which r are in information places. The probability of choosing a code at random which will fail to correctly decode such an error pattern is less than or equal to

$$P(j,r)=\sum_{i=0}^{j}\left[\binom{n}{i}-\binom{n-k}{i-r}\right]2^{-(n-k)} \qquad (18)$$

Each error pattern of this kind has probability $p^j q^{n-j}$ and there are $\binom{k}{r}\binom{n-k}{j-r}$ of them. Summing over j and r gives the over-all bound on average probability of error

$$P_{av}'=\sum_{j=0}^{n}\sum_{r=0}^{j}P(j,r)\binom{k}{r}\binom{n-k}{j-r}p^j q^{n-j}$$

$$=\sum_{j=0}^{n}\sum_{r=0}^{j}\sum_{i=0}^{j}p^j q^{n-j}\binom{k}{r}\binom{n-k}{j-r}\times \left[\binom{n}{i}-\binom{n-k}{i-r}\right]2^{-(n-k)}$$

$$=2^{-(n-k)}\sum_{j=0}^{n}p^j q^{n-j}\left[\binom{n}{j}\sum_{i=0}^{j}\binom{n}{j}-\sum_{r=0}^{j}\sum_{i=0}^{j}\binom{n-k}{j-r}\binom{n-k}{i-r}\binom{k}{r}\right] \qquad (19)$$

The term for which $j=0$ vanishes, yielding equation 14. This can be refined somewhat by noting $P(j,r)$ is an upper bound on a probability and hence can be replaced by 1 if it is greater than 1. If this is done and if the negative terms in $P(j,r)$ are dropped, the expression for P_{av}' results.

Reference

1. See reference 2 of the paper.

Digital Calculation of Network Functions Used in Loss Formula Studies

R. W. FERGUSON
ASSOCIATE MEMBER AIEE

R. W. LONG
ASSOCIATE MEMBER AIEE

L. J. RINDT
NONMEMBER AIEE

A COMPLETE matrix of self and mutual drops, with respect to a single reference point, will completely describe the constraints of a network as far as the significant points are concerned. Thus, these drops are a useful tool in the solution of network problems and have long been recognized as such. Previous to the days of the a-c network calculator, they were in widespread use for the calculation of losses, regulation short-circuit currents, etc. With the advent of the loss formula,[1] these drops assumed a new role. Ironically the a-c network calculator, which once subjugated these drops to a relatively minor role, now permitted their economical use by providing a means of obtaining them by direct measurement. For some time now the network calculator has been used quite widely to determine these drops; however, this method still involved considerable time and patience, and there was a serious question as to whether the accuracy obtained was sufficient for all purposes.

These limitations prompted an investigation as to the feasibility of obtaining these drops using a digital computer. The result of this investigation was a digital program that calculated the self and mutual drops of a system with a high degree of accuracy and at the same time compared favorably with the a-c network calculator method cost. A natural by-product of this digital program was the elimination of the tedious labor involved in the other methods and the reduction of a-c network calculator time required for obtaining a loss formula.

During the course of use of this particular digital program, the problem of obtaining these drops for a large system as well as those for its subsystems arose. Although the same program could be used for each of the various systems, a possibility existed that the self and mutual drops of the various subsystems could be modified and combined to form the self and mutual drops of the entire system. The investigation of this problem led to a digital program which would combine the self and mutual drops of two or more systems to form the drops

---

Paper 58-890, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, New York, June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 28, 1958.

R. W. Ferguson, R. W. Long, and L. J. Rindt are with Westinghouse Electric Corporation, East Pittsburgh, Pa.

for the system composed of the combination of these subsystems. A ramification of this program is the inclusion of a new line or group of lines into the matrix of self and mutual drops for a basic system.

This paper considers the theoretical background of the afore-mentioned two digital programs. Since the heart of the digital program that is used to obtain the self and mutual drops from the system impedances is commonly used in digital calculations, no attempt will be made to show the details of the program. However, a block diagram of the digital program for combining self and mutual drops is included.

Obtaining Self and Mutual Drops from the System Impedances

The matrix of self and mutual drops as considered here will be a square symmetrical matrix with the number of rows and columns equal to one less than the number of busses of the system which it describes. Any term in the matrix may be defined as follows:

D_{nn} = the voltage drop from the reference bus to bus *n* with one per unit current entering the system at the reference bus and leaving at bus *n*

D_{nm} = the voltage drop from the reference bus to bus *m* with one per unit current entering the system at the reference bus and leaving at bus *n*

Having defined the objective, an analysis of the method of obtaining it follows.

Consider the sample network shown in Fig. 1, consisting of three busses and three lines. The currents I_2 and I_3 are considered as flowing out of the network, while I_1 is considered as flowing into the network. Bus *1* will be designated as the reference bus.

Kirchhoff's law states that the vector sum of all the currents flowing to any bus is equal to zero. Using this law and expressing the currents within a network in terms of the bus voltages and line admittances, equations can be written describing the currents entering and leaving the network. Equations 1, 2, and 3 describe these currents for the sample network.

$$(E_1-E_2)y_{1-2}+(E_1-E_3)y_{1-3}=I_1 \qquad (1)$$

$$(E_1-E_2)y_{1-2}+(E_3-E_2)y_{2-3}=I_2 \qquad (2)$$

$$(E_1-E_3)y_{1-3}-(E_3-E_2)y_{2-3}=I_3 \qquad (3)$$

Rearranging the terms of equations 1, 2, and 3 the familiar form of the nodal equation is

$$E_1(y_{1-2}+y_{1-3})+E_2(-y_{1-2})+E_3(-y_{1-3})=I_1 \qquad (4)$$

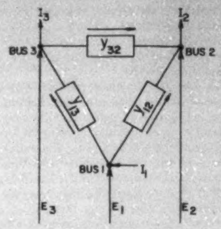


Fig. 1. Simple network showing pertinent parameters

$$E_1(-y_{1-2})+E_2(y_{1-2}+y_{2-3})+E_3(-y_{2-3})=-I_2 \qquad (5)$$

$$E_1(-y_{1-3})+E_2(-y_{2-3})+E_3(y_{1-3}+y_{2-3})=-I_3 \qquad (6)$$

These equations express the fact that the currents flowing into the network at a bus are equal to the sum of the products of *Ey*, where the self-admittance term, or the diagonal admittance of the particular row of a matrix, is obtained by summing the mutual admittance, or off diagonal terms, while the remaining *y*'s are the line admittance terms involved with their sign changed.

Since the self and mutual drops concern themselves with voltage drops from the reference bus, and since bus *1* is designated as the reference bus in the sample system, the voltage at bus *1* is the reference voltage. This means that the reference voltage may be written as equal to the voltage at bus *1* as in equation 7, and the voltage drops from the reference bus to bus *2* and *3* as in equations 8 and 9, respectively.

$$E_R=E_1 \qquad (7)$$

$$V_2=E_R-E_2 \qquad (8)$$

$$V_3=E_R-E_3 \qquad (9)$$

Combining equations 4 through 9 results in

$$E_R(y_{1-2}+y_{1-3})+(E_R-V_2)(-y_{1-2})+(E_R-V_3)(-y_{1-3})=I_1 \qquad (10)$$

$$E_R(-y_{1-2})+(E_R-V_2)(y_{1-2}+y_{2-3})+(E_R-V_3)(-y_{2-3})=-I_2 \qquad (11)$$

$$E_R(-y_{1-3})+(E_R-V_2)(-y_{2-3})+(E_R-V_3)(y_{1-3}+y_{2-3})=-I_3 \qquad (12)$$

An examination of these equations shows that all of the E_R terms drop out and the resultant equations appear as

$$V_2(-y_{1-2})+V_3(-y_{1-3})=-I_1 \qquad (13)$$

$$V_2(y_{1-2}+y_{2-3})+V_3(-y_{2-3})=I_2 \qquad (14)$$

$$V_2(-y_{2-3})+V_3(y_{1-3}+y_{2-3})=I_3 \qquad (15)$$

Since the sum of the currents into the network must equal the sum of the currents out of the network,

$$I_1=I_2+I_3 \qquad (16)$$

With this relationship established, equation 13 can be eliminated, since equations 13, 14, and 15 do not form a group of independent equations. A further simplification can be made by recognizing that the admittance terms are of the form of driving point and transfer admittances.[3] Thus, equations 14 and 15 can be rewritten as

$$V_2Y_{22}+V_3Y_{23}=I_2 \qquad (17)$$

$$V_2Y_{23}+V_3Y_{33}=I_3 \qquad (18)$$

These equations may then be solved to obtain the voltage drops from the reference bus (V_2 and V_3) in terms of the driving point and transfer admittances and the currents I_2 and I_3.

Now reconsider the definitions previously stated for D_{nn} and D_{nm} and write the equations to fit these definitions. First, if a unit current enters at bus *1* and leaves at bus *2*, $I_2=1$ and $I_3=0$. Then $D_{22}=V_2$ and $D_{23}=V_3$ can be obtained from the solution of

$$V_2Y_{22}+V_3Y_{23}=1 \qquad (19)$$

$$V_2Y_{23}+V_3Y_{33}=0 \qquad (20)$$

If a unit current is now allowed to enter at bus *1* and leave at bus *3*, $I_2=0$ and $I_3=1$. $D_{33}=V_3$ and $D_{32}=V_2$ can then be obtained from the solution of

$$V_2Y_{22}+V_3Y_{23}=0 \qquad (21)$$

$$V_2Y_{23}+V_3Y_{33}=1 \qquad (22)$$

The solution of these two sets of equations are readily seen to involve just those operations which one uses to find the inverse of a matrix. That is, both sets may be solved at the same time by considering the problem of finding the *D*'s in the matrix equation 23, which is expressed in equations 19 through 22 in matrix form, recognizing that the voltage drops represent the self and mutual drops.

$$\left\|\begin{matrix}D_{22}D_{23}\\D_{32}D_{33}\end{matrix}\right\|\cdot\left\|\begin{matrix}Y_{22}Y_{23}\\Y_{23}Y_{33}\end{matrix}\right\|=\left\|\begin{matrix}1&0\\0&1\end{matrix}\right\| \qquad (23)$$

Thus the conclusion that the self and mutual drop matrix (*D* matrix) is just the inverse of the driving point and transfer admittance matrix (*Y* matrix) with its first row and column deleted.

Digital Program

This is the problem that has been programmed for the IBM *704*. Each bus of a given power system is numbered in order, making sure that the reference bus is assigned *1*. All lines connected to bus *1* are listed with its impedances and then

Table I. Form for Tabulating Input Data*

Row	Column	Real Impedance	Imaginary Impedance
1	2	R_{12}	X_{12}
1	3	R_{13}	X_{13}
2	3	R_{23}	X_{23}

* Used for program which calculates self and mutual drops.

Table II. Typical Times for Inverting Admittance Matrix

Number of Busses	Time in Minutes
22	5
48	25
81	63

all those not already listed which are connected to bus *2*, etc. In Fig. 1 they would be listed as shown in Table I. Each row of this table would represent the information punched on one card. These cards are fed into the computer along with a card denoting the number of busses and lines. The program automatically proceeds to form the driving point and transfer admittance matrix; then deletes its first row and column; and inverts the matrix.

The program as outlined here is a complex double-precision matrix inversion program and has been used successfully in inverting several different systems. Examples of the size of the matrix and the time for inversion is shown in Table II. Naturally these times reflect the fact that there are only a limited number of lines connected to any one bus and that they are applicable only to typical power system problems.

Combination of Self and Mutual Drop Matrices

The problem to be considered in this section is the development of the self and mutual drops for a system from the self and mutual drops for two or more subsystems which make up this system. These subsystems are connected together at certain tie points and it is upon this fact that the solution is developed.

ASSUMPTIONS

The equations to be developed are based on the following assumptions:

1. The reference bus in each subsystem is chosen so as not to coincide with any of the tie points between them.
2. When the tie points of a system are listed in ascending order of bus numbers, they will agree exactly with a similar list made using the bus numbers of these tie points in the connecting system. An example of this would be as follows: If busses *5*, *7*, and *10* are the tie points of one system and busses *8*, *12*, and *14* are the tie points of the connecting system, then busses *5* and *8*, *7* and *12*, and *10* and *14* should refer to the same points respectively.

Assumption 2 makes the bookkeeping in the problem a little easier and negates the necessity of interchanging certain columns and rows of the matrices involved.

NOMENCLATURE

Assume a power system made up of two subsystems which are represented as system *A* and system *B*. Such a system is shown in Fig. 2. Systems *A* and *B* are connected to each other at certain tie points. The self and mutual drops for the two systems are represented by the *D* matrices in the equations

$$\|E^A\| = \|D^A\| \cdot \|I^A\| \tag{24}$$

$$\|E^B\| = \|D^B\| \cdot \|I^B\| \tag{25}$$

where the superscript *A* and *B* are used to refer to the *A* and *B* systems respectively.

To clarify the nomenclature further, each bus will be defined as one of the following three types: (1) reference bus designated by subscript *R;* (2) tie bus designated by the subscript *T;* (3) all other designated by the subscript *G*. When these subscripts are used for matrices and submatrices, they designate all such rows and columns of a system matrix. Consider the self and mutual drop matrix for system *A*.

$$\left\|\begin{array}{llll} D_{11}{}^A & D_{12}{}^A & D_{13}{}^A \ldots\ldots\ldots & D_{1T}{}^A \\ & D_{1(T+1)}{}^A \ldots\ldots\ldots & D_{1N}{}^A & \\ D_{21}{}^A & D_{22}{}^A & D_{23}{}^A \ldots\ldots\ldots & D_{2T}{}^A \\ & D_{2(T+2)}{}^A \ldots\ldots\ldots & D_{2N}{}^A & \\ \vdots & & & \\ D_{T1}{}^A & D_{T2}{}^A & D_{T3}{}^A \ldots\ldots\ldots & D_{TT}{}^A \\ & D_{T(T+1)}{}^A \ldots\ldots\ldots & D_{TN}{}^A & \\ D_{(T+1)1}{}^A & D_{(T+1)2}{}^A & D_{(T+1)3}{}^A \ldots & D_{(T+1)T}{}^A \\ & D_{(T+1)T(T+1)}{}^A \ldots\ldots & D_{(T+1)N}{}^A & \\ \vdots & & & \\ D_{N1}{}^A & D_{N2}{}^A & D_{N3}{}^A \ldots\ldots\ldots & D_{NT}{}^A \\ & D_{N(T+1)}{}^A \ldots\ldots\ldots & D_{NN}{}^A & \end{array}\right\| \tag{26}$$

The subscript *T* is used to indicate those terms involving tie points. The submatrix, which is formed by those terms having *T* in both the column and row subscript, is designated by $D_{TT}{}^A$. To simplify the subscript nomenclature further, the subscript *P* will designate all the rows or columns of the self and mutual drop matrix. Thus the entire matrix of the self and mutual drops will be referred to as $D_{PP}{}^A$ and the complete tie column submatrix as $D_{PT}{}^A$. The submatrices $D_{PP}{}^A$, $D_{PT}{}^A$, and $D_{TT}{}^A$ are shown diagrammatically as in Figs. 3(A), 3(B), and 3(C), respectively. It is to be seen the matrix $D_{PP}{}^A$ contains all of the terms of $D_{PT}{}^A$ and $D_{TT}{}^A$ and that $D_{PT}{}^A$ contains all of the terms of $D_{TT}{}^A$. These matrices were purposely defined this way because each tie point will appear as a point in the combined system.

Since the tie points will only appear once in the combined system matrix and the *A* system matrices were defined to include them, the *B* system matrices will be defined without them. Thus $D_{GG}{}^B$, $D_{GT}{}^B$ and $D_{TT}{}^B$ are mutually exclusive and so have no terms in common. These three matrices are shown schematically in Fig. 4(A), 4(B), and 4(C) respectively.

The use of a single subscript will denote a column vector consisting of all the row terms that are designated by the subscripts. For example, $I_T{}^A$ is the column vector of the currents at the tie busses. $I_T{}^{A\prime}$ will have the customary meaning of being the transpose of $I_T{}^A$, which in this case would be a row vector with a similar definition. Another notation of this type would be $E_R{}^A - E_P{}^A$. This will be taken to mean the column vector of the voltage drops from the reference bus in the *A* system to all other busses in the *A* system. Similarly, $E_R{}^A - E_R{}^B$ will be taken as the voltage drop from the reference bus of the *A* system to the reference bus of the *B* system.

Lastly, it will be well to point out that all of the terms of the various matrices and submatrices are complex numbers and complex arithmetic is involved in these calculations. Submatrices designated as 1 and 0 are thus $1+JO$ and $0+JO$ respectively. While the 1 submatrices may be superscripted and subscripted in accordance with the notation above, the 0 submatrices will be assumed to have the proper number of rows and columns to make the matrices involved conformable.

DEVELOPMENT OF THE EQUATIONS

Using the preceding notation, matrix equations can be written for the *A* and *B* systems respectively

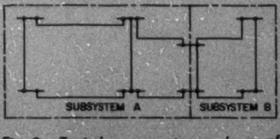


Fig. 2. Typical power system made up of two subsystems

$$\left\|\begin{array}{c} E_R{}^A-E_P{}^A \\ \hdashline E_R{}^A-E_T{}^A \end{array}\right\| = \left\|\begin{array}{c:c} D_{PP}{}^A & D_{PT}{}^A \\ \hdashline D_{TP}{}^A & D_{TT}{}^A \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_T{}^A \end{array}\right\| \qquad (27)$$

$$\left\|\begin{array}{c} E_R{}^B-E_G{}^B \\ \hdashline E_R{}^B-E_T{}^B \end{array}\right\| = \left\|\begin{array}{c:c} D_{GG}{}^B & D_{GT}{}^B \\ \hdashline D_{TG}{}^B & D_{TT}{}^B \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline I_T{}^B \end{array}\right\| \qquad (28)$$

In these two equations, it is understood that the various quantities $D_{PP}{}^A$, $I_P{}^A$, $E_R{}^B-E_G{}^B$, etc., are submatrices having the appropriate number of rows and columns.

Each term of the left hand matrix of equation 28 may be typically rewritten

$$(E_R{}^B-E_N{}^B)=(E_R{}^A-E_N{}^B)-(E_R{}^A-E_R{}^B) \qquad (29)$$

This permits equation 28 to be written

$$\left\|\begin{array}{c} E_R{}^A-E_G{}^B \\ \hdashline E_R{}^A-E_T{}^B \end{array}\right\| - (E_R{}^A-E_R{}^B)\left\|\begin{array}{c} 1_G{}^B \\ \hdashline 1_T{}^B \end{array}\right\| = \left\|\begin{array}{c:c} D_{GG}{}^B & D_{GT}{}^B \\ \hdashline D_{TG}{}^B & D_{TT}{}^B \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline I_T{}^B \end{array}\right\| \qquad (30)$$

Since the sum of all the currents entering and leaving system B must be equal to zero,

$$0=\left\|\begin{array}{c:c:c} 1_G{}^{B'} & 1_T{}^{B'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline I_T{}^B \\ \hdashline I_R{}^B \end{array}\right\| \qquad (31)$$

As each tie current in system B is the negative of this same current in system A,

$$\|I_T{}^B\|=\|-I_T{}^A\| \qquad (32)$$

This relationship may be substituted into equations 30 and 31 and they may be then combined to give

$$\left\|\begin{array}{c} E_R{}^A-E_G{}^B \\ \hdashline E_R{}^A-E_T{}^B \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| - (E_R{}^A-E_R{}^B)\left\|\begin{array}{c} 1_G{}^B \\ \hdashline 1_T{}^B \\ \hdashline 1 \end{array}\right\| = \left\|\begin{array}{c:c:c} D_{GG}{}^B & D_{GT}{}^B & 0 \\ \hdashline D_{TG}{}^B & D_{TT}{}^B & 0 \\ \hdashline 1_G{}^{B'} & 1_T{}^{B'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline -I_T{}^A \\ \hdashline I_R{}^B \end{array}\right\| \qquad (33)$$

Consider the second row of equations 27 and 33. They are respectively

$$\|E_R{}^A-E_T{}^A\|=\left\|\begin{array}{c:c} D_{TP}{}^A & D_{TT}{}^A \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_T{}^A \end{array}\right\| \qquad (34)$$

$$\|E_R{}^A-E_T{}^B\|=(E_R{}^A-E_R{}^B)\|1_T{}^B\|+ \left\|\begin{array}{c:c:c} D_{TG}{}^B & D_{TT}{}^B & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline -I_T{}^A \\ \hdashline I_R{}^B \end{array}\right\| \qquad (35)$$

Since the tie points in each system are identical points, when the systems are combined the voltages at the tie points of each system are the same. This gives

$$\|E_R{}^A-E_T{}^A\|=\|E_R{}^A-E_T{}^B\| \qquad (36)$$

Having this relationship, the right hand members of equations 34 and 35 may be equated

$$(E_R{}^A-E_R{}^B)\|1_T{}^B\|=\left\|\begin{array}{c:c} D_{TP}{}^A & D_{TT}{}^A \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_T{}^A \end{array}\right\| - \left\|\begin{array}{c:c:c} D_{TG}{}^B & D_{TT}{}^B & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_G{}^B \\ \hdashline -I_T{}^A \\ \hdashline I_R{}^B \end{array}\right\| \qquad (37)$$

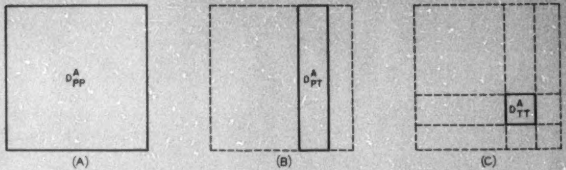


Fig. 3. Various submatrices of the A system's self and mutual drop matrix

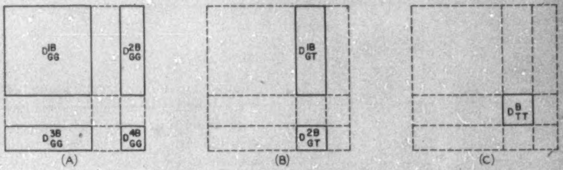


Fig. 4. Various submatrices of the B system's self and mutual drop matrix

$$\text{A}\text{—}D_{GG}{}^B=\left\|\begin{array}{c:c} D_{GG}{}^{1B} & D_{GG}{}^{2B} \\ \hdashline D_{GG}{}^{3B} & D_{GG}{}^{4B} \end{array}\right\|$$

$$\text{B}\text{—}D_{GT}{}^B=\left\|\begin{array}{c} D_{GT}{}^{1B} \\ \hdashline D_{GT}{}^{2B} \end{array}\right\|$$

Rearranging the terms in equation 37 results in

$$(E_R{}^A-E_R{}^B)\|1_T{}^B\|=\left\|\begin{array}{c:c:c} D_{TP}{}^A & -D_{TG}{}^B & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| + \|D_{TT}{}^A+D_{TT}{}^B\| \cdot \|I_T{}^A\| \qquad (38)$$

The third row of equation 33 can be rewritten as

$$0=\left\|\begin{array}{c:c:c} 0 & 1_G{}^{B'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| - \|1_T{}^{B'}\| \cdot \|I_T{}^A\| \qquad (39)$$

Although the form of equations 38 and 39 is not necessarily the simplest, the current matrix represents the form needed in the solution of the problem.

Since the number of ties is the same in each system, $1_T{}^B$ is the same as $1_T{}^A$. This enables the combining of equations 38 and 39 into

$$\left\|\begin{array}{c:c} -D_{TT}{}^A-D_{TT}{}^B & 1_T{}^A \\ \hdashline 1_T{}^{A'} & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_T{}^A \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| = \left\|\begin{array}{c:c:c} D_{TP}{}^A & -D_{TG}{}^B & 0 \\ \hdashline 0 & 1_G{}^{B'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| \qquad (40)$$

The first matrix in equation 40 can be eliminated from the left hand side by multiplying both sides of the equations by the inverse of this matrix. If the product of the inverse matrix and the third matrix of equation 40 is designated as matrix C, then equation 40 can be written as

$$\left\|\begin{array}{c} I_T{}^A \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| = \|C\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| \qquad (41)$$

Equation 41 will be substituted into the appropriate places of equations 27 and 33. That is, in the first row of equation 27 and also in the first and third row of equation 33. To make this process a little easier to visualize, these rows will be rewritten as equations 42, 43, and 44, respectively.

$$\|E_R{}^A-E_P{}^A\|=\left\|\begin{array}{c:c:c} D_{PP}{}^A & 0 & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| + \left\|\begin{array}{c:c} D_{PT}{}^A & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_T{}^A \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| \qquad (42)$$

$$\|E_R{}^A-E_G{}^B\|=\left\|\begin{array}{c:c:c} 0 & D_{GG}{}^B & 0 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| + \left\|\begin{array}{c:c} -D_{GT}{}^B & 1_G{}^B \end{array}\right\| \cdot \left\|\begin{array}{c} I_T{}^A \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| \qquad (43)$$

$$\|E_R{}^A-E_R{}^B\|=\left\|\begin{array}{c:c:c} 0 & 1_G{}^{B'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_P{}^A \\ \hdashline I_G{}^B \\ \hdashline I_R{}^B \end{array}\right\| + \left\|\begin{array}{c:c} -1_T{}^{A'} & 1 \end{array}\right\| \cdot \left\|\begin{array}{c} I_T{}^A \\ \hdashline E_R{}^A-E_R{}^B \end{array}\right\| \qquad (44)$$

By substituting equation 41 into equations 42, 43, and 44 and combining the three resultant equations into one matrix equation, an equation is produced which can be solved to obtain one row of the self and mutual drop matrix for the

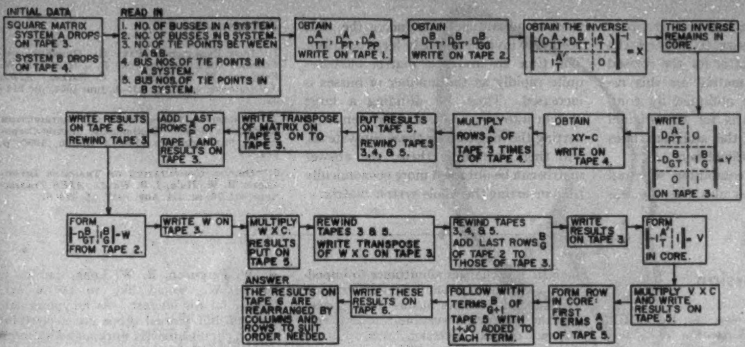

INITIAL DATA
SQUARE MATRIX SYSTEM A DROPS ON TAPE 3.
SYSTEM B DROPS ON TAPE 4.
READ IN
1. NO. OF BUSSES IN A SYSTEM.
2. NO. OF BUSSES IN B SYSTEM.
3. NO. OF TIE POINTS BETWEEN A & B.
4. BUS NOS. OF TIE POINTS IN A SYSTEM.
5. BUS NOS. OF TIE POINTS IN B SYSTEM.
OBTAIN $D_{TT}^A, D_{PT}^A, D_{PP}^A$ WRITE ON TAPE 1.
OBTAIN $D_{TT}^B, D_{GT}^B, D_{GG}^B$ WRITE ON TAPE 2.
OBTAIN THE INVERSE $\left\| \begin{array}{c:c} (D_{TT}^A + D_{TT}^B & 1_T^A) \\ \hdashline 1_T^{A'} & 0 \end{array} \right\|^{-1} = X$
THIS INVERSE REMAINS IN CORE.
WRITE $\left\| \begin{array}{c:c} D_{PT}^A & 0 \\ \hdashline -D_{GT}^B & 1_G^B \\ \hdashline 0 & 1 \end{array} \right\| = Y$ ON TAPE 3.
OBTAIN XY=C WRITE ON TAPE 4.
MULTIPLY ROWS $_P^A$ OF TAPE 3 TIMES C OF TAPE 4.
PUT RESULTS ON TAPE 5. REWIND TAPES 3, 4, & 5.
WRITE TRANSPOSE OF MATRIX ON TAPE 5 ON TO TAPE 3.
ADD LAST ROWS $_P^A$ OF TAPE 1 AND RESULTS ON TAPE 3.
WRITE RESULTS ON TAPE 6. REWIND TAPE 3
FORM $\left\| -D_{GT}^B \,\vdots\, 1_G^B \right\| = W$ FROM TAPE 2.
WRITE W ON TAPE 3.
MULTIPLY W X C. RESULTS PUT ON TAPE 5.
REWIND TAPES 3 & 5. WRITE TRANSPOSE OF W X C ON TAPE 3.
REWIND TAPES 3, 4, & 5. ADD LAST ROWS $_G^B$ OF TAPE 2 TO THOSE OF TAPE 3.
WRITE RESULTS ON TAPE 3.
FORM $\left\| -1_T^{A'} \,\vdots\, 1 \right\| = V$ IN CORE.
MULTIPLY V X C AND WRITE RESULTS ON TAPE 5.
FORM ROW IN CORE: FIRST TERMS $_G^A$ OF TAPE 5.
FOLLOW WITH TERMS $_{G+1}^B$ OF TAPE 5 WITH 1+J0 ADDED TO EACH TERM.
WRITE THESE RESULTS ON TAPE 6.
ANSWER
THE RESULTS ON TAPE 6 ARE REARRANGED BY COLUMNS AND ROWS TO SUIT ORDER NEEDED.

Fig. 5 (left). Block diagram of program for combining the self and mutual drop matrix for system A and B. (Superscript and subscript notation is the same as used in the text)

combined A and B system. This equation is shown as

$$\left\| \begin{array}{c} E_R^A - E_P^A \\ \hdashline E_R^A - E_G^B \\ \hdashline E_R^A - E_R^B \end{array} \right\| = \left\{ \left\| \begin{array}{c:c:c} D_{PP}^A & 0 & 0 \\ \hdashline 0 & D_{GG}^A & 0 \\ \hdashline 0 & 1_G^{B'} & 1 \end{array} \right\| + \left\| \begin{array}{c:c} D_{PT}^A & 0 \\ \hdashline -D_{GT}^B & 1_G^B \\ \hdashline -1_T^{A'} & 1 \end{array} \right\| \cdot \|C\| \right\} \cdot \left\| \begin{array}{c} I_P^A \\ \hdashline I_G^B \\ \hdashline I_R^B \end{array} \right\| \quad (45)$$

To solve equation 45 for a complete matrix of self and mutual drops would require that it be solved for as many cases as there are rows in the current matrix. In each case all of the currents except one would be zero, the remaining current would be one. If all of the equations for these different cases were written as one matrix equation, then the current matrix would be a unit matrix of appropriate size and the resulting equation would be

$$\|D\| = \left\| \begin{array}{c:c:c} D_{PP}^A & 0 & 0 \\ \hdashline 0 & D_{GG}^B & 0 \\ \hdashline 0 & 1_G^{B'} & 1 \end{array} \right\| + \left\| \begin{array}{cc} D_{PT}^A & 0 \\ -D_{GT}^B & 1_G^B \\ -1_T^{A'} & 1 \end{array} \right\| \cdot \|C\| \quad (46)$$

where the D matrix is the self and mutual drop matrix for the system formed by combining system A and B.

It should be pointed out here that to solve equation 46 requires matrix multiplication and addition with one matrix inverse. The inverse of a matrix is used to determine matrix C. However, this is a small square matrix which size is equal to the number of tie points plus one.

Digital Program

An IBM *704* program has been written to solve equation 46. A general-type block diagram is shown in Fig. 5 illustrating the steps followed by the program in obtaining the combined matrix of self and mutual drop matrix for an over-all system from self and mutual drop matrix of two systems which combined form the over-all system.

This program has been applied to systems of 37 and 94 busses respectively. The first system was formed by two subsystems while the second was formed from three subsystems. In the latter case, two of the subsystem matrices were combined and the resultant matrix was combined with that of the third subsystem.

The Addition of a Line to a System

A variation of the matrix combining program would be the case where a system is changed by adding a line or a small network of lines. If more than one line is added, the network of these lines would be treated as a small system. The self and mutual drop coefficients for this small system could be calculated and then combined with the main system as outlined in the previous sections. However, the case of adding a single line resolves into a special case where the form of the equations is greatly simplified.

Consider this case of adding one line. To obtain this set of self and mutual drops, the original system is taken as subsystem A. A reference bus is now arbitrarily placed somewhere along this new line, and the two points where this line meets subsystem A are designated as tie points. The branch impedance of the new line may be proportioned in some manner to establish the impedance between the fictitious reference bus and the two tie points. This will establish subsystem B, consisting of the reference bus and two tie points. The self and mutual drop are now obtained for this system and consists of a two-by-two matrix. For this matrix D_{GG}^B and D_{GT}^B are zero so that the new equation corresponding to equation 44 is shown as equation 47.

$$\left\| \begin{array}{c} E_R^A - E_P^A \\ \hdashline E_R^A - E_G^B \\ \hdashline E_R^A - E_R^B \end{array} \right\| = \left\{ \left\| \begin{array}{c:c:c} D_{PP}^A & 0 & 0 \\ \hdashline 0 & 0 & 0 \\ \hdashline 0 & 0 & 1 \end{array} \right\| + \left\| \begin{array}{c:c} D_{PT}^A & 0 \\ \hdashline 0 & 0 \\ \hdashline -1_T^{A'} & 1 \end{array} \right\| \cdot \|C\| \right\} \cdot \left\| \begin{array}{c} I_P^A \\ \hdashline I_G^B \\ \hdashline I_R^B \end{array} \right\| \quad (47)$$

Since the row and column involving the reference bus of the subsystem B will not be needed in the final matrix for this case, the last row and column from this matrix can be deleted. Further, since the second row and column are zero everywhere, they can also be deleted. Thus, the calculation of the final self and mutual drop matrix for the case of adding a single line to a system for which the drop matrix is already known can be simplified. The matrix C as described previously is obtained and then its last row and column are deleted. Thus equation 47 can be reduced to

$$\|E_R^A - E_P^A\| = \|D_{PP}^A\| + \|D_{PT}^A\| \cdot \|C^*\| \cdot \|I_P^A\| \quad (48)$$

where the matrix C^* is the same as matrix C with its last row and column deleted.

Equation 48 now can be expanded to represent the entire self and mutual drop matrix.

$$\|D\| = \|D_{PP}^A\| + \|D_{PT}^A\| \cdot \|C^*\| \quad (49)$$

Conclusions

Although this method of obtaining the self and mutual drop matrix justifies itself economically, the elimination of tedious a-c network calculator measurements and the increased accuracy makes this method desirable from many standpoints.

When the self and mutual drop matrix is known for a system and the system is changed by adding a line or a group of lines, it is not necessary to reinvert the entire admittance matrix for the revised system. The drop matrix for this revised system can be obtained by combining the drop matrix for the original system with that of the additions lines using the method outlined here.

This method for combining self and mutual drop matrices also has the less obvious advantage of reducing the time needed to invert extremely large matrices. The time to invert a matrix increases quite rapidly as the number of busses is increased. Thus, by dividing a large system into two or three subsystems, inverting the matrices for these subsystems, and then combining these, the answer matrix can be obtained more economically than inverting the whole system matrix.

References

1. LOSS EVALUATION, PART II. CURRENT- AND POWER-FORM LOSS FORMULAS, E. L. Harder, R. W. Ferguson, W. E. Jacobs, D. C. Harker. *AIEE Transactions*, vol. 73, pt. III-A, June 1954, pp. 716–31.

2. ELECTRICAL TRANSMISSION AND DISTRIBUTION REFERENCE BOOK. Westinghouse Electric Corporation, Pittsburgh, Pa., fourth edition, 1950, p. 314.

3. DIGITAL COMPUTATION OF TRANSFER IMPEDANCES, H. W. Hale, J. B. Ward. *AIEE Transactions*, vol. 76, pt. III, Aug. 1957, pp. 476–81.

Discussion

E. E. George (Ebasco Services, Inc., Little Rock, Ark.): This derivation of a loss formula for an interconnected system is particularly welcome because it is adapted to direct use on an IBM *704* or a Remington-Rand *1103A*, or a similar large digital computer (smaller systems could be handled on a smaller type of computer).

The authors' method is not too difficult to follow, it is simple matrix algebra. Their method eliminates construction of a tree diagram or other manual analysis heretofore required.

The inversion of a nodal admittance matrix, after removing the reference bus row and column, is a powerful tool in circuit analysis. It changes admittance to impedance, converts a mesh to a star network with the old reference bus as the new neutral, the process is reversible, and as is well known, matrix inversion solves linear simultaneous equations in a convenient form.

Matrix algebra badly needs convenient symbols for the new matrix formed by adding or deleting a reference bus row and column to or from a previously designated matrix. While the authors follow the European use of a prime symbol as a superscript to denote transposition, some of us prefer the subscript t to denote transposition.

Also the asterisk symbol is generally used in mathematics and engineering to denote a conjugate.

R. W. Ferguson, R. W. Long, and L. J. Rindt: We would like to thank Mr. George for his interest. As he points out, the self and mutual drops are a powerful tool in circuit analysis work and with more and more of this type of work being programmed for high-speed digital computers, the use of these drops will undoubtedly increase considerably.

While the authors agree that some standard set of symbols are needed for matrix designations, the absence of standard symbols led them to use symbols that would not conflict with the various superscripts and subscripts of the submatrices, used in this particular development. It is hoped that with the growing use of matrix algebra, some standard set of symbols will be adopted.

Digital Computer Aids in Power-Pool-Operation Studies

H. M. McINTYRE
ASSOCIATE MEMBER AIEE

C. W. BLAKE
ASSOCIATE MEMBER AIEE

J. S. CLUBB
NONMEMBER AIEE

Synopsis: This paper describes the use of a general-purpose digital computer as an aid in the solution of problems encountered in the pooled operation of a large hydroelectric system. This program has been used effectively during the past year by operations engineers of the Bonneville Power Administration (BPA). Its use has enabled studies of operations to be prepared more rapidly and in a form that can be used readily by management in analyzing current conditions and in formulating plans for subsequent operations.

THE 11 major utilities in the Pacific Northwest have been operating as a power pool since early 1940. The utilities operate in the states of Oregon, Washington, Idaho, Montana, and Utah, and in British Columbia, Canada. The resources in the area are predominantly hydroelectric and much is gained through co-ordination of the interconnected pool and nonpool generating facilities. The ability to take advantage of stream flow and load diversity has increased firm load-carrying capability by an estimated 1,000 mw (megawatts).

The installed capacity of the pool, including amounts available from interconnected nonpool utilities, is more than 9,000 mw of hydroelectric capacity and more than 1,000 mw of thermal generating facilities. Projects currently under construction or definitely scheduled will add 7,000 mw of hydroelectric capacity and 1,100 mw of thermal generating capacity. About 63%, or 10,100 mw, of the total hydroelectric facilities will be concentrated in a series of 17 projects. About 7,110 mw will be installed at the federal projects in this series. These projects are shown in Fig. 1, and start with Hungry Horse in western Montana on the upper reaches of south fork of the Flathead River, 1,150 miles from the Pacific Ocean at an elevation of 3,560 feet above mean sea level. The last of this series is the Bonneville plant only 150 miles from the Pacific Ocean, located at tidewater on the lower Columbia River. Several of the projects are not in operation but are either under construction or definitely scheduled. Projects under construction are John Day, Priest Rapids, and Rocky Reach. The Wanapum project is scheduled for construction after completion of Priest Rapids.

The importance of a co-ordinated operation on the series of 17 plants can be understood better when it is realized that they represent six different ownerships, private, public, and federal. The significance to this series of projects of pool operation is further emphasized when it

---

Paper 58-892, recommended by the AIEE Computing Devices Committee and the System Engineering Committee of the Power Apparatus and Systems Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing May 15, 1958.

H. M. MCINTYRE, C. W. BLAKE, and J. S. CLUBB are with the Bonneville Power Administration, Portland, Oreg.

The authors wish to express their indebtedness to L. A. Dean and others of the Bonneville Power Administration for their work in preparing the plant data and flow charts for the program described in this paper.

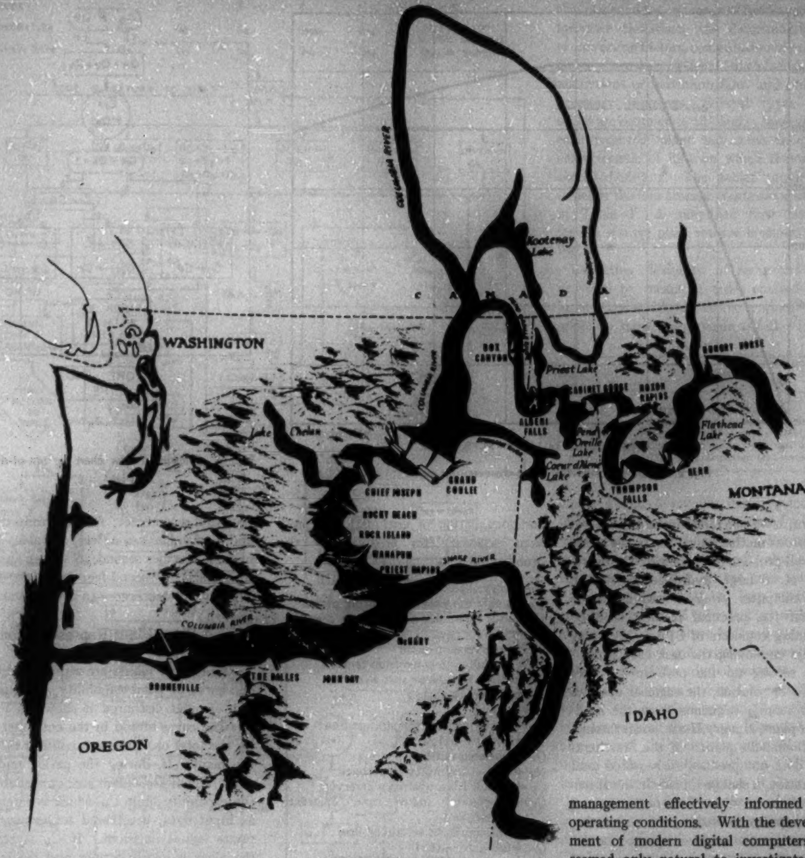

COLUMBIA RIVER
Kootenay Lake
KOOTENAY RIVER
C A N A D A
WASHINGTON
BOX CANYON
Priest Lake
HUNGRY HORSE
CABINET GORGE
NOXON RAPIDS
COLUMBIA RIVER
ALBENI FALLS
Flathead Lake
Lake Chelan
Pend Oreille Lake
SPOKANE RIVER
Coeur d'Alene Lake
KERR
GRAND COULEE
CHIEF JOSEPH
THOMPSON FALLS
MONTANA
ROCKY REACH
ROCK ISLAND
WANAPUM
PRIEST RAPIDS
SNAKE RIVER
COLUMBIA RIVER
McNARY
BONNEVILLE
THE DALLES
JOHN DAY
IDAHO
OREGON

Fig. 1. Major series of hydroelectric projects in the Columbia River Basin

is realized that out of a total reservoir capacity of almost 20 million acre-feet, 67% is usable through all or part of this series of 17 plants. The large storage reservoirs and the magnitude of installed capacity make it possible for this series of projects to act as the leveling factor, or the balance wheel, for the entire pool. Grand Coulee is the key project, having some 5 million acre-feet of usable storage and 2,212 mw of generating capacity at full operating load.

A co-ordinating group is retained by power-pool members to develop seasonal analysis, collect and maintain current operating records, and to prepare in advance a pooled operating program for the subsequent year under critical and median month stream-flow conditions. It is necessary, however, for BPA to develop a pool-type analysis for plants shown in Fig. 1 in order to operate the federal system most effectively within the pool framework and yet to maintain the system's ability to meet all contractual commitments.

This requires a continuing analysis of pool operations by BPA as loads and resources diverge from original assumptions. The continuing expansion of the system has made it increasingly difficult to complete operation studies in time to keep management effectively informed on operating conditions. With the development of modern digital computers, it seemed only natural to investigate the possibility of developing a machine solution of the system operations problem. The program described in this paper has been written for the Bendix *G-15A* computer, although similar programs have been used on the International Business Machines Corporation *650* and the ALWAC IIIE computers.

The Problem

The problem is essentially one of setting up an efficient schedule of generation for the system of hydroelectric plants in order to meet an estimated load. Schedules are set up on a daily, weekly, monthly, and seasonal basis. The multi-decisions necessary in sequence, each one interdependent upon the previous de-

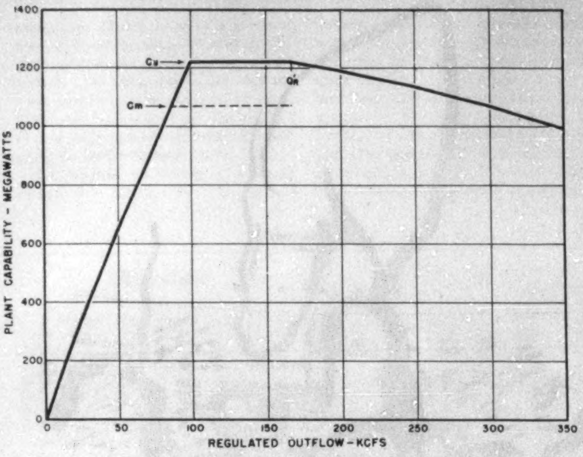


Fig. 2. Typical run-of-river plant characteristic curve

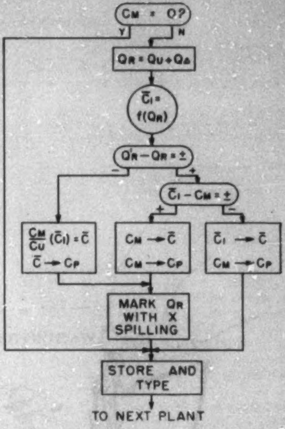


Fig. 3. Simplified flow chart for run-of-river plant

cisions, have precluded up to the present a rigorous mathematical treatment of the over-all problem of optimizing the operation of a large hydroelectric system. The computer program was developed around the practical method used by operating engineers of BPA in preparing studies employing the desk calculator.

In setting up the problem for desk calculator solution, the engineer operates the system in sequence, from the headwater plant, Hungry Horse, down through the Bonneville plant. If the first iteration does not produce the required total generation, a change is made in the operation of one or more of the reservoirs and the computation made again. In setting up changes, the engineer must be careful not to violate any physical constraints of the system.

Plant output and reservoir data are obtained from plant characteristic curves which have been put in tabular form to facilitate reading and to allow consistent interpolation between values when using a desk calculator. In the computer program these tables are stored in the memory and operated upon by means of table look-up and linear, single-interpolation subroutines. The tables have been reduced to as few points as is required to define the original curves within reasonable accuracy.

Nomenclature

The following symbols have been adopted in order to facilitate flow-charting and formulation of arithmetical expressions necessary for programming the problem.

$\bar{C}$ = energy capability
C_M = maximum capability during period
C_P = peak capability
C_U = ultimate capability limit
H/K = conversion factor
H/K_{FG} = full-gate conversion factor
F = forebay elevation
T = tailwater elevation
H = head
Q_D = flow resulting from storage draft
Q_M = modified inflow
Q_R = regulated outflow
Q_R' = flow at beginning of cutback
Q_U = outflow from upstream reservoir
Q_Δ = incremental inflow from upstream reservoir
Q_{max} = maximum or restricted flow
Q_{FG} = full-gate outflow
S = storage content
S_i = initial storage content
S_e = end storage content
S_Δ = change in storage content
t = length of period, days

Capability values are in megawatts; conversion factors are in megawatts per thousand second-feet; forebay, tailwater, and head values are in feet; Q values are in cubic feet per second; S values are in thousands of second-feet-days. A bar over a symbol indicates an average value.

Run-of-River Plant Representation

Run-of-river plants are represented by one curve of plant capability as a function of plant discharge as shown in Fig. 2. The curve stored in the computer is for the total or ultimate capability C_U of the plant. If, during the period under study, all units of a plant are not available, the capability limit C_M, which is entered as input data, is adhered to by appropriate logical decisions. If C_M is zero, the plant is not in service and the program skips to the next plant. The point at which rising flows cause plant capabilities to drop off because of reduced head is designated as the cutback point Q_R' in Fig. 2, and the capability will follow the curve or be proportioned down-

Table I. Data Required to Define Operation of Storage Plants

Reservoir and Plant Characteristics	Reservoir			
	Hungry Horse	Flathead Lake	Albeni Falls	Grand Coulee
$Q_{FG}=f(H)$	X	X	X	X
$H/K=f(H)$	X	X	X	X
$Q_{max}=f(S)$	X	X	X	X
$Q_{FG}=f(S)$	X	X	X	X
$S=f(F)$	X	X		X
$T=f(Q_R)$	X	X		X
$H/K_{FG}=f(H)$				X
$H=f(S \text{ and } Q_R)$			X	
$C=f(Q)$ low flows	X			

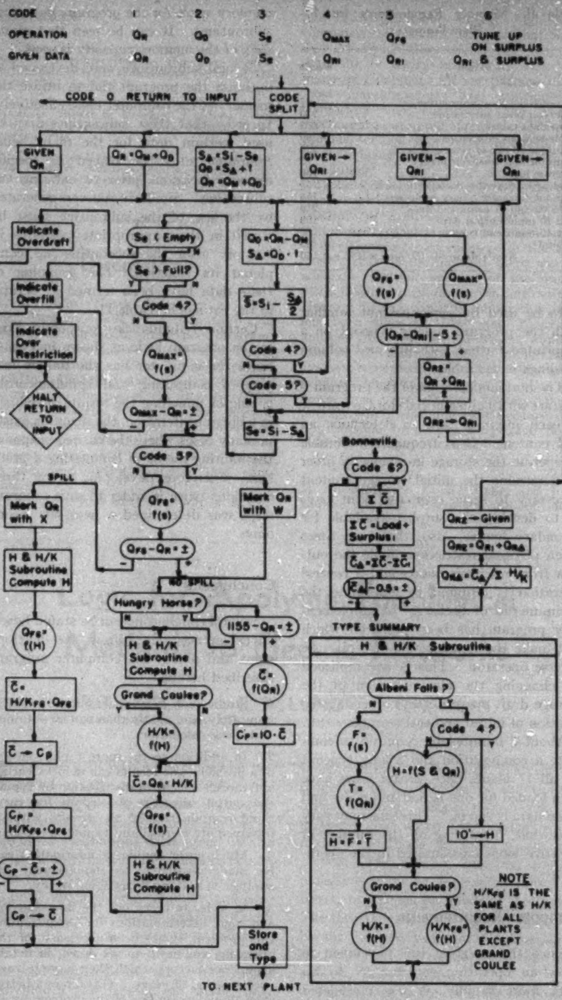

CODE 1 2 3 4 5 6
OPERATION Q_R Q_D S_e Q_{MAX} Q_{FG} TUNE UP ON SURPLUS
GIVEN DATA Q_R Q_D S_e Q_{RI} Q_{RI} Q_{RI} & SURPLUS
CODE 0 RETURN TO INPUT
CODE SPLIT
GIVEN Q_R
$Q_R = Q_M + Q_D$
$S_\Delta = S_I - S_e$
$Q_D = S_\Delta \div 1$
$Q_R = Q_M + Q_D$
GIVEN → Q_{RI}
GIVEN → Q_{RI}
GIVEN → Q_{RI}
Indicate Overdraft
S_e < Empty
Y
N
S_e > Full?
Y
N
Indicate Overfill
Code 4?
N
Y
Indicate Over Restriction
$Q_{MAX} = f(s)$
HALT RETURN TO INPUT
$Q_{MAX} - Q_R = \pm$
−
+
Code 5?
N
Y
$Q_D = Q_R - Q_M$
$S_\Delta = Q_D \cdot 1$
$\bar{S} = S_I - \frac{S_\Delta}{2}$
Code 4?
N
Y
Code 5?
N
Y
$S_e = S_I - S_\Delta$
$Q_{FG} = f(s)$
$Q_{MAX} = f(s)$
$|Q_R - Q_{RI}| - 5 \pm$
−
+
$Q_{RZ} = \frac{Q_R + Q_{RI}}{2}$
$Q_{RZ} \rightarrow Q_{RI}$
Bonneville
Code 6?
N
Y
$\Sigma \bar{C}$
$\Sigma \bar{C}$ = Load + Surplus$_1$
$\bar{C}_\Delta = \Sigma\bar{C} - \Sigma\bar{C}_1$
$|\bar{C}_\Delta| - 0.5 = \pm$
−
+
TYPE SUMMARY
Q_{RZ} → Given
$Q_{RZ} = Q_{RI} + Q_{R\Delta}$
$Q_{R\Delta} = \bar{C}_\Delta / \Sigma\, H/K$
SPILL
Mark Q_R with X
H & H/K Subroutine Compute H
$Q_{FG} = f(H)$
$\bar{C} = H/K_{FG} \cdot Q_{FG}$
$\bar{C} \rightarrow C_p$
$C_p = H/K_{FG} \cdot Q_{FG}$
$C_p - \bar{C} = \pm$
−
+
$C_p \rightarrow \bar{C}$
$Q_{FG} = f(s)$
Mark Q_R with W
$Q_{FG} - Q_R = \pm$
−
+
NO SPILL
Hungry Horse?
N
Y
$1155 - Q_R = \pm$
−
+
H & H/K Subroutine Compute H
Grand Coulee?
Y
N
$H/K = f(H)$
$\bar{C} = Q_R \cdot H/K$
$Q_{FG} = f(s)$
H & H/K Subroutine Compute H
$\bar{C} = f(Q_R)$
$C_P = 10 \cdot \bar{C}$
Store and Type
TO NEXT PLANT
H & H/K Subroutine
Albeni Falls?
N
Y
$F = f(s)$
$T = f(Q_R)$
$\bar{H} = \bar{F} - \bar{T}$
Code 4?
N
Y
$H = f(S \,\&\, Q_R)$
10′ → H
Grand Coulee?
N
Y
$H/K = f(H)$
$H/K_{FG} = f(H)$
NOTE
H/K_{FG} IS THE SAME AS H/K FOR ALL PLANTS EXCEPT GRAND COULEE

Fig. 4. Simplified flow chart for storage plant

ward by the ratio of C_M to C_U if the flow exceeds this point. The plant data for each run-of-river plant are based on an average forebay elevation. No allowance in this program is made for operation of pondage at run-of-river plants due to the excessive memory space requirements.

A simplified flow chart for a run-of-river plant is shown in Fig. 3.

Storage Plant Representation

A storage plant is represented in the computer memory by several tables of characteristics which allow the computation of a head for any range of forebay and tailwater elevations, the determination of conversion factors, computation of peak and average output, tests for violation of restrictions or maximum flow, and other necessary functions. Several types of tables are common to all plants; however, each reservoir plant has some unique characteristic of its own which requires special handling. The tables required to define the four storage plants are shown in Table I. A simplified flow chart for the storage plant routine is shown in Fig. 4.

Operating flexibility of reservoirs is provided by means of codes entered as input data which allows the engineer to select the mode of operation as follows:

Code	Given Data
1	regulated outflow
2	flow resulting from storage draft
3	end of period storage content
4	restricted outflow (estimated)
5	full-gate discharge (estimated)

Codes 4 and 5 involve iterative loops which use the estimated given data for the first iteration. Each successive iteration uses the average of the first value and the computed value as the new given data and pulls out of the loop when a convergence test of 5 cubic feet per second is met. Generally three to five iterations are required for convergence depending on the value of the original estimate.

Code 6, as shown in Fig. 4, is provided to tune the complete system to a given load by varying the storage operation of any one of the four reservoirs at the option of the operating engineer. When using this code the operation of three of the reservoirs is fixed and the computer is instructed to meet the total system load by operation of the remaining reservoir. The load is actually given to the computer as the difference between the generation and the estimated total load. This difference is given as zero if the total load is to be met exactly. The logic in this operation is to provide a convenient way of computing steam replacement (if the difference is positive) or of dropping the nonfirm load (if the difference is negative). This operation includes within the overall iterative process, the iterative loops of codes 4 and 5. Convergence is quite rapid, however, because the conversion factors used in the solution are developed by the program operation. This process makes the operation quite stable in most cases and converges monotonically in two to three iterations. Oscillitory convergence results when spill is occurring at some plants below the reservoir which is being used to take the swing. In this case as many as ten iterations may be required.

In the event of an overdraft, an overfill, or the violation of a reservoir outlet restriction resulting from an invalid operating instruction, the computer types out an indicator of the violation along with a reservoir identification code, and then halts. The engineer must then specify an operation that will allow the computation to proceed.

Program Operation

Operation of the program for one period requires input data for each plant consisting of a stream flow and a maximum capability limit. In addition, the initial storage of each reservoir must be specified. In each subsequent period the initial storage content is carried over from the end content of the preceding period.

Because of the magnitude of the program, it was found necessary to store all the stream-flow and maximum capability-limit data on punched paper tape. These data are read into the computer via the photoreader for one period. The operating code and given data for each storage plant are entered via the typewriter for the first period and the computation started. Either a complete (long) or an abbreviated (short) typeout can be obtained as the computation proceeds. The usual method of operation consists of setting up the reservoir operation criteria for each period and running through the complete study, obtaining a short typeout for each period as it is computed. When this is completed, the results are reviewed and any necessary adjustments are made in the system operation by the operator in any period which seems to need modification. After the complete operation has been made, a long typeout covering all plants is made for each period.

Originally it was intended that very detailed data on each plant and reservoir would be included in the output format. This was given up, however, because of the memory capacity required just to store these data waiting for typeout. The typeout for run-of-river plants consists only of peak and average capability and plant discharge, while for a storage plant values for S_Δ, S_e, and H/K are typed out in addition. The short typeout previously mentioned includes only the storage plant data and a summary giving the total generation, load, and surplus. In many cases it has been found that the short typeout of a complete study is ample information; however, if the study is to be used by personnel not familiar with the program, a long typeout on a preprinted format with line and column headings is desirable.

Table II. Memory Requirements for 17-Plant Program

Item	Words
Run-of-river plant subroutine	47
Reservoir plant subroutine	229
Plant data tables	602
Table look-up subroutine	80
Setup for 13 run-of-river plants	196
Setup for four reservoir plants	73
Program logic	136
Loading in (includes decimal-to-binary)	155
Typeout (includes binary-to-decimal)	155
Lines 00 and 01 command lines	216
Line 19 input-output data	108
Line 15 input and computed data	108
Total	2,105
Available words (20 lines × 108) =	2,160

The normal operation of the program is to start with a given initial storage content of each reservoir and to determine an end content. It is frequently desirable to operate the storage in reverse in order to determine the initial storage content necessary to serve estimated firm loads or to develop the surplus available for secondary loads. Also, there are times when physical restrictions limit the outflow from reservoir plants and a reverse operation is required to determine the optimum rate of discharge. A supplementary program has been prepared which will make the necessary changes for the reverse operaton. This is accomplished by changing the algebraic sign of the storage draft operation and reversing the position of initial and end storage.

About 3 minutes are required to complete a computation and a long typeout for all 17 plants once stream flows have been loaded for one period and codes and given data typed in. This time may vary somewhat, depending on the number of iterative loops encountered in the operation.

Computer Requirements

Since the program was developed to allow an operating engineer to do his work more rapidly, it was considered essential that decimal-to-binary and binary-to-decimal subroutines be retained in the memory at all times to enable the engineer to "talk" to the computer without acquiring an intimate knowledge of what goes on inside the computer. This requirement, of course, uses memory capacity, but enables an engineer to learn the essentials of the program and to adapt himself to working with the computer quite rapidly.

Table II shows the allocation of memory space for the program, data, and subroutines. It can be seen that about 98% of the memory capacity is used.

Several subroutines were developed to optimize the program and to utilize the memory of the computer more effectively. In order that these subroutines may be used, certain data for the plant being computed must be placed in specific memory locations prior to entering the subroutine. Similarly, answers generated by the use of the subroutine must be stored in their appropriate positions in memory after the subroutine has completed its function. The grouping of these data have been termed a "setup" in the listings in Table II.

Certain redundancies purposely exist in the program logic as shown in Fig. 4. Since the computer has the unique feature of computing and simultaneously typing or punching the results developed in a previous iteration, the situation exists in many cases where the desired output is the limiting factor in completing a problem. Therefore, developing two or three functions twice in order to save program steps was determined a necessary economy.

Conclusions

Several conclusions can be stated based on experience gained from the development and use of the computer program described herein.

1. Studies can be prepared more rapidly, accurately, and neatly than can be obtained with desk calculators.
2. In addition to the more rapid solution of a problem, the results can achieve earlier and more widespread distribution by typing the output directly on vellum for easy, rapid reproduction of the results without intermediate clerical and typing operations.
3. Much more significant information can be made available to management for analysis in making operational policy.
4. In order to carry out the flow-charting and logic determinations for programming this problem, it has been necessary for the operating engineers to set down, in detail, experience factors which they have learned over a period of years. This forces a better over-all understanding of the problem by all personnel.

Continued research appears warranted in order to increase the effectiveness of the present program. The use of auxiliary storage devices is being investigated in order to enlarge the scope of the program. Research on an analytical approach to plant operation is being carried out in an effort to reduce the memory requirements for plant representation and to decrease the computation time.

Vol.
Pt. 1
195

References

1. Computer Search for Economical Operation of a Hydrothermal Electric System, R. J. Cypser. *AIEE Transactions*, vol. 73, pt. III-B, Oct. 1954, pp. 1260-67.

2. Computational Aspects of Dynamic Programming, Stuart E. Dreyfus. *Pamphlet P-924*, The Rand Corporation, Santa Monica, Calif., Nov. 19, 1956.

3. Theory of Dynamic Programming, R. Bellman. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 60, no. 6, Nov. 1954, p. 503.

◆

Discussion

G. W. Bills (North American Aviation, Inc., Downey, Calif.): It is gratifying to read of the successful power-pool programming with the aid of the Bendix *G-15A* digital computer. When I was employed at BPA in 1955 there was considerable speculation as to the use of the Bendix *G-15A* on this problem because of the possible accuracy limitation of the computer and programming difficulties. What is your current appraisal of the use of the Bendix as compared with the ALWAC IIIE computer?

Obviously your next step will be to employ a large-memory-storage high-speed computer such as the International Business Machines Corporation *709* together with the necessary transducers and telemetering equipment for the purpose of almost instantaneously and continuously computing and controlling the Pacific Northwest power-pool operation. This will provide for the short-range program, but what is being planned for the long-range scheduling?

In 1954, if I remember correctly, the contract work being done at Massachusetts Institute of Technology for BPA on the optimization of water flow and minimization of electric line losses ended with the conclusion that at the moment there was no adequate solution because of the unpredictability of stream flows over long periods. Has any other attempt at solving this problem been made by BPA since then? It appears that with the increasing application of stochastic processes to industry, stochastic methods might be employed to produce long-range scheduling of water power in a more effective fashion, on the average, than assuming that because it is not possible to predict long-range flows, there is nothing that can be done.

H. M. McIntyre, C. W. Blake, and **J. S. Clubb:** Mr. Bill's discussion and continued interest in system operation problems are appreciated.

Programming difficulties of the *G-15A* have been overcome by thorough training in machine language programming. It has been necessary, however, for the machine language programmer to become more or less specialized, which would not necessarily be the case with a computer employing single-address-command system such as the ALWAC IIIE.

A program for use with an International Business Machines Corporation *704* is presently being written which embodies all of the features of the *G-15A* program plus control routines which eliminate operator intervention in the preparation of seasonal operating studies.

Since the termination of the research contract between BPA and Massachusetts Institute of Technology considerable effort has been spent by BPA on the development of long-range stream-flow forecasts. A semistatistical method of determining a "dependable" and a "probable" stream-flow estimate from 1 to 7 months in advance has been used quite successfully during the past 2 years. This method provides stream-flow forecasts which are used as input data for operation studies. This approach also provides a convenient method of determining flow forecasts of varying probability.

Logic for Applying Topological Methods to Electric Networks

R. T. BYERLY
ASSOCIATE MEMBER AIEE

R. W. LONG
ASSOCIATE MEMBER AIEE

C. W. KING
NONMEMBER AIEE

THE DETERMINATION of the behavior of large electric networks confronts the engineer or mathematician with the necessity of solving a large number of equilibrium equations determined by the network and the restraints imposed on it. This task has been sufficiently formidable to bring about the development and construction of large analog computers such as the d-c and a-c network calculators. In recent years, the availability of medium- and large-scale digital computers has provided a new facility for solving network problems. Considerable attention has been given to applying digital computers to load flow, transient stability, and short-circuit calculations.

For the most part, those engaged in this work have relied on nodal analysis methods because of the relative ease of establishing network relationships in the computer from basic topological and impedance data.

However, there are reasons to prefer in some cases to use the equilibrium equations obtained with mesh analysis. A reduction in the number of equations is often obtained. More stable coefficient matrices may speed convergence where iterative methods of solution are employed. The ease of handling mutual inductive coupling especially raises interest in the mesh analysis approach. The formation of mesh current equations by the analyst is of course a possibility, but this involves considerable time and effort for large systems.

These motives have led the authors to explore the possibility of forming mesh current relationships from raw system data using a large-scale digital computer. As will be seen shortly, the heart of the matter is concerned with the generation of a matrix which contains in usable form the information regarding network configuration which is needed to establish the required number of independent mesh equations. The logic required to obtain such a matrix is described for two different approaches. One method constructs the matrix a column at a time, the other constructs the matrix a row at a time.

Review of Some Geometrical Aspects of Networks

As an introduction to this problem, it would be well for the reader to review parts of the book by Professor Ernst A. Guillemin.[1] In the first chapter the motivation and the introduction to the application of topological methods to network analysis are quite clearly presented.

In developing a procedure for finding network relations by topological methods, we are solely interested in the network geometry, and therefore represent all elements as lines with small circles at the ends to designate the terminals. Such a representation is called a graph of the network. The graphs with which we will be concerned consist of these lines which

---

Paper 58-826, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 14, 1958.

R. T. Byerly, R. W. Long, and C. W. King are with Westinghouse Electric Corporation, East Pittsburgh, Pa.

The authors wish to express their appreciation to C. C. Li for his assistance in the early phases of this study. Mr. Li is a graduate student at Northwestern University and was employed in the Analytical Department of Westinghouse at East Pittsburgh during the summer of 1957.

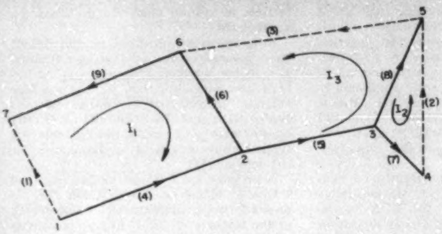


Fig. 1. Network no. 1 showing links, tree branches, and loops

we call branches and the small circles which we refer to as nodes. Using the nomenclature of Guillemin, we refer to the total number of branches of the network as b and the total number of nodes as n_t. This totality of branches for the network sets up a number of closed paths over which currents can circulate. It is apparent that by removing certain branches this property for carrying current will be destroyed. In Fig. 1 there is represented a network graph with $b=9$ and $n_t=7$. By removing the dotted lines, we destroy its current carrying property. These dotted lines are referrred to as links and the remainder of the network is referred to as the tree.

The tree has the particular property that it contains no closed paths. The lines which make up a tree will be called tree branches while the dotted lines are link branches. It is evident that for our given network graph several different trees could be found, however, the number of tree branches for each tree will always be equal to $n_t-1=6$. The number of link branches is therefore $b-6=3$. These formulas stem from the fact that a tree is said to consist of any set of branches in the given graph which is just sufficient to connect all the nodes. This is truly just n_t-1, for the addition of one more line would join two nodes which must necessarily have been connected and thus would create a closed path.

Professor Guillemin has shown that the values of the L link currents, where $L=b-(n_t-1)$, must fix all the currents of the network. That is, all the tree branch currents may be uniquely expressed in terms of the link currents. Therefore, there are only L independent currents in a netowrk and L is the smallest number of currents in terms of which all the other may be uniquely expressed.

We may thus uniquely express the state of the network in terms of L independent variables. Since each link causes a loop in the network to close, the current through that link can be used to identify one of the loop currents for the network. The correspondence between link and loop currents is shown in Fig. 1. Positive currents are assumed to flow from the low numbered node to the high numbered node in each branch. The branch currents are related to the loop or link currents in Fig. 1 by the following set of equations:

$I_1=I_1$

$I_2=I_2$

$I_3=I_3$

$I_4=-I_1$

$I_5=+I_3$

$I_6=-I_1-I_3$

$I_7=I_2$

$I_8=-I_2+I_3$

$I_9=-I_1$

In matrix notation:

$$\begin{Vmatrix} I_1\\ I_2\\ I_3\\ I_4\\ I_5\\ I_6\\ I_7\\ I_8\\ I_9 \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ -1 & 0 & 0\\ 0 & 0 & 1\\ -1 & 0 & -1\\ 0 & 1 & 0\\ 0 & -1 & 1\\ -1 & 0 & 0 \end{Vmatrix} \begin{Vmatrix} I_1\\ I_2\\ I_3 \end{Vmatrix}$$

Now designate the voltage drops (IZ) in the nine branches as:

$V_1,\ V_2,\ V_3,\ V_4,\ V_5,\ V_6,\ V_7,\ V_8,\ V_9$

and similarly the generation in the nine branches as

$E_1,\ E_2,\ E_3,\ F_4,\ E_5,\ E_6,\ E_7,\ E_8,\ E_9$

and note the Kirchhoff's law of voltages applied to each of the three loops gives the set of equations

$V_1-V_9-V_6-V_4=E_1-E_9-E_6-E_4$

$V_2-V_8+V_7=E_2-E_8+E_7$

$V_3-V_6+V_5+V_8=E_3-E_6+E_5+E_8$

In matrix notation:

$$\begin{Vmatrix} 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1\\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0 \end{Vmatrix} \begin{Vmatrix} V_1\\ V_2\\ V_3\\ V_4\\ V_5\\ V_6\\ V_7\\ V_8\\ V_9 \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1\\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0 \end{Vmatrix} \begin{Vmatrix} E_1\\ E_2\\ E_3\\ E_4\\ E_5\\ E_6\\ E_7\\ E_8\\ E_9 \end{Vmatrix}$$

If the coefficient matrix immediately above is designated T, then the coefficient matrix in the current relationship is T', the transpose of T. In general:

$TV_B=TE_B$

$I_B=T'I_L$

Now form an impedance matrix for the network for Fig. 1 as follows:

$$Z_B=\begin{Vmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} & Z_{15} & Z_{16} & Z_{17} & Z_{18} & Z_{19}\\ Z_{21} & Z_{22} & Z_{23} & Z_{24} & Z_{25} & Z_{26} & Z_{27} & Z_{28} & Z_{29}\\ Z_{31} & Z_{32} & Z_{33} & Z_{34} & Z_{35} & Z_{36} & Z_{37} & Z_{38} & Z_{39}\\ Z_{41} & Z_{42} & Z_{43} & Z_{44} & Z_{45} & Z_{46} & Z_{47} & Z_{48} & Z_{49}\\ Z_{51} & Z_{52} & Z_{53} & Z_{54} & Z_{55} & Z_{56} & Z_{57} & Z_{58} & Z_{59}\\ Z_{61} & Z_{62} & Z_{63} & Z_{64} & Z_{65} & Z_{66} & Z_{67} & Z_{68} & Z_{69}\\ Z_{71} & Z_{72} & Z_{73} & Z_{74} & Z_{75} & Z_{76} & Z_{77} & Z_{78} & Z_{79}\\ Z_{81} & Z_{82} & Z_{83} & Z_{84} & Z_{85} & Z_{86} & Z_{87} & Z_{88} & Z_{89}\\ Z_{91} & Z_{92} & Z_{93} & Z_{94} & Z_{95} & Z_{96} & Z_{97} & Z_{98} & Z_{99} \end{Vmatrix}$$

In this array, the diagonal terms are the branch impedances and the off diagonal terms are the inductive mutual impedances. One may write:

$V_B=Z_BI_B$

Multiplying on the left by T,

$TV_B=TZ_BI_B$

Substituting for I_B,

$TV_B=TZ_BT'I_L$

or finally,

$TE_B=(TZ_BT')I_L$

This last matrix equation is the set of mesh current equations required for the solution of a network. To establish this set of equations it is not only necessary to create a method for obtaining the T matrix but it is important that the method not be too cumbersome and unwieldy. The remainder of this paper describes two different approaches to this problem and the solution to which each approach has led.

A Method for Determining the Link Branches and the Tree Branches for a Given Network

The procedure to be followed in writing a transformation matrix T by columns consists of two parts: the first part is concerned with the separation of the network

Table I

Line Nos.	Branches By Node Nos.	
12	1	4
1	1	6
13	1	9
16	1	9
14	1	10
10	2	4
5	2	5
2	2	6
17	2	6
3	2	7
9	2	10
7	3	8
8	3	10
11	4	9
4	5	7
6	5	8
15	7	10

Table II

Link Nos.	Node Nos. R.	C.
1	1	9
2	4	9
3	2	6
4	2	6
5	2	10
6	2	7
7	5	7
8	5	8

Table III

Tree Branch Nos.	Node Nos. R.	C.
9	1	4
10	1	6
11	1	9
12	1	10
13	2	4
14	3	10
15	7	10
16	2	5
17	3	8

Table IV

✓	✓	✓	✓	✓	✓	✓	✓	✓
4	6	9	10	2	3	7	5	8

branches into tables of link branches and tree branches and the second part concerns the formation of the matrix from these tables. The method of selection of links and tree branches is the same regardless of the complexity of the network and the same digital computer program will solve the problem for all cases.

As can be well appreciated, some orderly way of considering each branch of the network must be employed. This is accomplished through the use of a table which contains a listing of branches by numbers along with a designation of terminal node numbers for each branch. A separation of the branches into links and tree branches is based on data contained in such a table. The concepts which must be kept in mind in analyzing this data are best illustrated by considering a specific example.

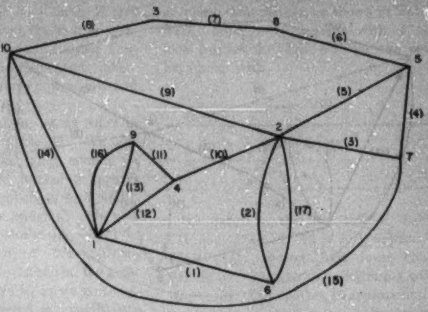


Fig. 2. Network graph for network 2

Consider the network shown in Fig. 2. Assume that each branch of the network has been given a line number from Table I which is a listing of all the branches of the network. Each branch is listed just once with the lower numbered node being listed first and the higher numbered node second. This in effect will set up the positive direction for every branch as we will consider the positive direction to be from the lower numbered node to the higher numbered node. Further it should be noted that in listing the branches of Table I, we first write down all the branches emanating from node 1, then all those from node 2, and then 3, etc.

Having formed Table I we prepare to form Tables II–V. Table II is the listing of link branches, Table III is to contain tree branches, Table IV is an auxiliary device which is used to establish the order in which successive nodes are examined, and Table V is a correspondence between original line numbers and the numbers assigned to links and tree branches in Tables II and III. Begin by saying that all branches through node 1 are tree branches and must be listed in Table III unless there are parallel branches. Since there are parallel branches between nodes 1 and 9, the second 1 9 branch closes a loop so it must be listed as a link branch in Table II. In Table IV we now place the numbers 4, 6, 9, and 10 which are the new nodes we meet in moving along the foregoing branches.

We now consider the first new node we have encountered, namely 4, which is listed first in Table IV. We go back to Table I, ignoring node 1 since branches through node 1 have already been accounted for, and hunt for all branches through 4. The first one we find is the branch 2 4. Since node 2 has not been used before, branch 10, which connects nodes 2 and 4 cannot possibly close a loop. It is therefore entered in Table III as a tree branch the node 2 is placed in Table IV just to the right of node 10. The next branch is 4 9. Remember that node 9 appears to the right of 4 in Table IV and therefore we have previously used node 9; 4 9 must close a loop so it is a link and should be entered in Table II. There are no more 4's in Table I, so we never have to consider 4 again. This is shown in Table IV by putting a check above the 4.

The next node in Table IV is 6, so we will return to Table I to find all branches through node 6. Again we ignore the 1's. The first branch containing a 6 is 2 6 but the number 2 appears to the right of 6 in Table IV; therefore, both nodes 2 and 6 have been used before so 2 6 is entered as a link in Table II. The next 6 is encountered as 2 6 which is a parallel branch to the first 2 6 branch therefore, it too must be entered into Table II as a link. There are no more 6's so that a check is put in the first line of Table IV above the 6.

The next number appearing in Table IV is 9 so we must find all 9's in Table I. The first 9 appears in 4 9. Now since 4 appears to the left of 9 in Table IV, branch 4 9 has been considered before so should now be passed over. We have now exhausted the 9's, so we check off 9 in Table IV and pick up the next number appearing there.

We must now find all 10's appearing in Table I. We first encounter 2 10; 2 appears to the right of 10 in Table IV,

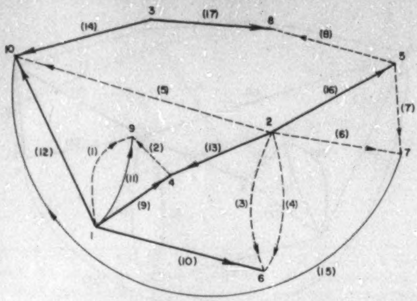

3
(17)
8
(8)
(14)
10
5
(5)
(16)
(7)
2
(6)
7
9
(12)
(1)
(2)
(13)
(11)
4
(9)
(3)
(4)
1
(10)
6
(15)

Fig. 3. Link and tree branch graph for network 2

therefore 2 10 must be entered in Table II as a link. The next 10 in Table I appears as 3 10, since node 3 does not appear in Table IV, it is written in Table IV and the branch 3 10 is entered into Table III as a tree branch. The next 10 appears in Table I as 7 10; the 7 does not appear in Table IV so it is written in Table IV and the branch 7 10 entered into Table III as a tree branch. This exhausts the 10's so a check is placed above the 10 in Table IV.

In the foregoing manner we now consider all the 2's in Table I, then all the 3's all the 7's, the 5's, and the 8's. Table II now contains all the link branches and Table III all the tree branches.

We now fill in the first column of these tables starting with Table II and the number 1 and reassign new line numbers to these links and tree branches. The new line numbers facilitate the formation of the transformation matrix T by columns. Table V is now formed as a coordinating table which sets up a correspondence between the new branch numbers and those appearing in Table I.

The general flow of logic used in the procedure for obtaining the link and tree branches is shown with the block diagram in Fig. 4. The given network is redrawn in Fig. 3. In this figure, the tree branches are shown as solid lines and the link branches as dotted, the arbitrary directions are shown and each line is numbered according to the new assignment shown in Table V.

Table V

Original Branch No.	New Branch No.
16	1
11	2
2	3
17	4
9	5
3	6
4	7
6	8
12	9
1	10
13	11
14	12
10	13
8	14
15	15
5	16
7	17

FORMATION OF THE TRANSFORMATION MATRIX T BY COLUMNS

Consider again the T matrix previously written for the network of Fig. 1. Write branch numbers above the columns and link numbers before the rows as shown below:

$$\begin{array}{c|ccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
1 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1 \\
2 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
3 & 0 & 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0
\end{array}$$

Note that the columns of the T matrix may be regarded as relationships which determine each branch current in terms of link currents. The branches 1, 2, and 3 are links, hence the branch currents I_1, I_2, and I_3 are identical to link currents. The first three columns and rows constitute a 3 by 3 identity submatrix. In general, given n links, the T matrix formed using the procedure to be described will have as a submatrix an n by n identity matrix. The problem is thus reduced to forming the remaining columns of the matrix. The fundamental ideas needed for the formation of these remaining columns are:

1. Kirchhoff's Law which states that the sum of all currents at a node must equal zero.
2. Every branch current may be expressed in terms of one or more link currents.

Now let us again turn our attention to the network of Fig. 3, and examine node 4 of the network. If B_9 represents the current in branch 9, and $L2$ the current in link 2, we may state that

$$-B9-B13+L2=0$$

where the minus sign indicates current into the node and the plus sign indicates current out of the node. Further,

$$B9=-B13+L2 \qquad (1)$$

Before we can express $B9$ in terms of link currents, we must certainly express $B13$ in link currents and then substitute this value in equation 1. Upon examination of Fig. 3, we see that if we go to node 2 we may write

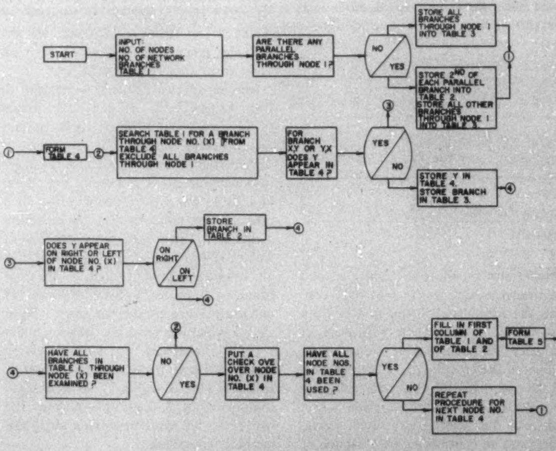

START
INPUT: NO. OF NODES NO. OF NETWORK BRANCHES TABLE I
ARE THERE ANY PARALLEL BRANCHES THROUGH NODE I ?
NO
YES
STORE ALL BRANCHES THROUGH NODE I INTO TABLE 3
STORE 2ND OF EACH PARALLEL BRANCH INTO TABLE 2. STORE ALL OTHER BRANCHES THROUGH NODE I INTO TABLE 3.
①
①
FORM TABLE 4
②
SEARCH TABLE I FOR A BRANCH THROUGH NODE NO. (X) FROM TABLE 4 EXCLUDE ALL BRANCHES THROUGH NODE I
FOR BRANCH X,Y OR Y,X DOES Y APPEAR IN TABLE 4 ?
YES
NO
③
STORE Y IN TABLE 4. STORE BRANCH IN TABLE 3.
④
③
DOES Y APPEAR ON RIGHT OR LEFT OF NODE NO. (X) IN TABLE 4 ?
ON RIGHT
ON LEFT
STORE BRANCH IN TABLE 2
④
④
④
HAVE ALL BRANCHES IN TABLE I, THROUGH NODE (X) BEEN EXAMINED ?
NO
YES
②
PUT A CHECK OVE OVER NODE NO. (X) IN TABLE 4
HAVE ALL NODE NOS. IN TABLE 4 BEEN USED ?
YES
NO
FILL IN FIRST COLUMN OF TABLE I AND OF TABLE 2
FORM TABLE 5
⑩
REPEAT PROCEDURE FOR NEXT NODE NO. IN TABLE 4
①

Fig. 4. Flow diagram for selecting links and tree branches

$$B13+L3+L4+L6+L5+B16=0 \qquad (2)$$

while node 5 enables us to write

$$-B16+L8+L7=0 \qquad (3)$$

Substitution from equations 2 and 3 into equation 1 gives

$$B9=L2+L3+L4+L5+L6+L7+L8$$

It is apparent that in order to express $B9$ in terms of links, it was necessary to include the definition of branches 13 and 16 in terms of links. We may think of branches 9, 13, and 16 as making up a chain of branches or a limb of our original tree. It is this notion that we wish to use and through its use develop a relatively easy method for expressing each branch current in terms of link currents. To accomplish this, we will next develop Tables VI and VII which list all nodes except node 1, and all links through each node in Table VI and all branches through each node in Table VII. Node 1 is omitted from Tables VI and VII because it is possible to establish only 9 independent branch current relationships for the network of Fig. 3.

Table VI

Node	Link
2	3, 4, 5, 6
3	
4	2
5	7, 8
6	−3, −4
7	−6, −7
8	−8
9	−1, −2
10	−5

Table VII

Node	Branch
2	13, 16
3	14, 17
4	−9, −13
5	−16
6	−10
7	15
8	−17
9	−11
10	−12, −14, −15

The general rule is the same for obtaining Table VI and Table VII. However, Table VI is obtained from using Table II while Table VII is obtained from Table III. The rule states that to fill in Tables VI for a given node, we first examine the right hand column of Table II and pick out all those places where this node occurs noting the link number for each of these places. Each of these link numbers is listed along side the node number in Table VI, with a plus sign being given the link number if the node number occurs under R in Table II and a minus sign being given the link number if the node number appears under C in Table II. This sign convention differentiates between currents flowing into and out of the node.

In way of illustration, let us list 2 in our node column of Table VI. We now proceed to the right hand column of Table II. Our node number 2 appears under R for link numbers 3, 4, 5, and 6 but does not appear under C. We now list these numbers along side our 2 in Table VI with plus signs. We next consider node 3, 4, etc., filling in the right hand column of Table VI as we go along. Table VII is completed in exactly this same manner, though we appreciate the fact that we must use Table III in order to pick out the necessary branches.

A few remarks on the makeup of Table VII would be in order. First we observe that each branch can appear no more than twice in the table and that if a branch appears twice it will appear with different signs as we are looking at it from each of the two nodes at which it terminates. Since there are only n_t-1 tree branches, we can say that there will be less than $2(n_t-1)$ numbers appearing in the second column of Table VII as node (1) does not appear in the table. Further, we see that every node listed in the left hand column of Table VII cannot be followed by two branch numbers in the right hand column. For if this were possible, we could get completely from one node back to itself again by means of tree branches and this would close a loop. We would therefore have an error in our Tables VI and VII. We can thus certainly say that there is at least one node listed in Table VII which has only one tree branch joined to it. The significance of this fact will become apparent as the process develops.

Before we proceed to the picking out our tree limbs, we can remark that since there are 8 links and 9 tree branches our T matrix will consist of 8 rows and 17 columns. Since branch 1 really means link 1 and branch 2 is link 2, etc., we see that the first 8 columns with the first 8 rows form an 8 by 8 identity matrix. So we are really concerned only with columns 9 through 17 of our T matrix at this time.

DESIGNATION OF LIMBS

Since column 9 of our matrix is the first one to be found, we look for the row of the right-hand column of Table VII which begins with 9 or −9. We see that 9 appears as −9 and has a −13 appearing with it. These two numbers are placed in a limb defining box as shown in Table VIII. Since −13 appears with 9 we look further in Table VII for the row beginning with 13.

Table VIII

Limb Branches		
−9	−13	16

It appears in the first row of Table VII along with 16; 16 is now written in Table VIII, and we look down Table VII till we find the row beginning with 16 or −16. The number 16 appears in the fourth row by itself so we have defined our limb which begins with branch 9. We now go to Table III and define our limb in terms of the nodes through which it passes. This is shown in Table IX.

Table IX

Limb Nodes					
1	4	2	4	2	5

Branch 9 passes through nodes 1 and 4, branch 13 through nodes 2 and 4, branch 16 through nodes 2 and 5, these are listed in Table IX. It is inherent to the process that the first node in the limb consisting of branches 9, 13, and 16 is not used in defining branch 9 in terms of links. The reader may verify this by reconsidering equations 1, 2, and 3. Hence, delete the first node number in Table IX. In this case it is a 1. We further delete the second 4 and second 2 which appear in Table IX, all repetitions of numbers are always deleted. This leaves 4, 2, and 5 as the nodes through which this limb passes. We now refer to Table VI under the node column and we write down the links which pass through these nodes. We obtain: link 2 passes through node 4, links 3, 4, 5, 6 pass through node 2, and links 7 and 8 pass through node 5. This defines branch 9 in terms of links and the ninth column of the T matrix appears in Table X. If one of the links had a minus sign, a −1 would have been put into the proper space of column 9.

Table X

	9	10	14
1	0	0	0
2	1	0	0
3	1	−1	0
4	1	−1	0
5	1	0	0
6	1	0	0
7	1	0	0
8	1	0	1

Table XI

Link No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Branch No.																
1	1	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	1	0	−1	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	1	−1	0	0	−1	0	0	0	0
4	0	0	0	1	0	0	0	0	1	−1	0	0	−1	0	0	0	0
5	0	0	0	0	1	0	0	0	1	0	0	−1	−1	0	0	0	0
6	0	0	0	0	0	1	0	0	1	0	0	−1	−1	0	1	0	0
7	0	0	0	0	0	0	1	0	1	0	0	−1	−1	0	1	1	0
8	0	0	0	0	0	0	0	1	1	0	0	−1	−1	1	0	1	−1

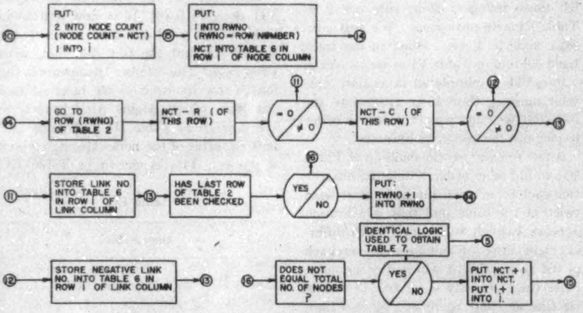

10
PUT: 2 INTO NODE COUNT (NODE COUNT = NCT) 1 INTO i
15
PUT: 1 INTO RWNO (RWNO = ROW NUMBER) NCT INTO TABLE 6 IN ROW i OF NODE COLUMN
14
14
GO TO ROW (RWNO) OF TABLE 2
NCT − R (OF THIS ROW)
= 0
≠ 0
11
NCT − C (OF THIS ROW)
= 0
≠ 0
12
13
11
STORE LINK NO INTO TABLE 6 IN ROW i OF LINK COLUMN
13
HAS LAST ROW OF TABLE 2 BEEN CHECKED
YES
NO
16
PUT: RWNO + 1 INTO RWNO
14
IDENTICAL LOGIC USED TO OBTAIN TABLE 7.
5
12
STORE NEGATIVE LINK NO. INTO TABLE 6 IN ROW i OF LINK COLUMN
13
16
DOES NOT EQUAL TOTAL NO. OF NODES ?
YES
NO
PUT NCT + 1 INTO NCT. PUT i + 1 INTO i.
15

Fig. 5. Flow diagrams for selecting links or tree branches through each node

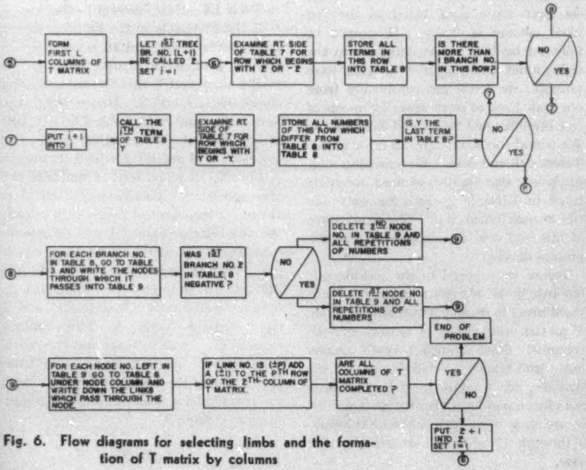

5
FORM FIRST L COLUMNS OF T MATRIX
LET 1ST TREE BR. NO. (L+1) BE CALLED Z. SET i = 1
6
EXAMINE RT. SIDE OF TABLE 7 FOR ROW WHICH BEGINS WITH Z OR −Z
STORE ALL TERMS IN THIS ROW INTO TABLE 8
IS THERE MORE THAN 1 BRANCH NO. IN THIS ROW?
NO
YES
8
7
7
PUT i + 1 INTO i
CALL THE iTH TERM OF TABLE 8 Y
EXAMINE RT. SIDE OF TABLE 7 FOR ROW WHICH BEGINS WITH Y OR −Y.
STORE ALL NUMBERS OF THIS ROW WHICH DIFFER FROM TABLE 8 INTO TABLE 8
IS Y THE LAST TERM IN TABLE 8?
NO
YES
7
8
8
FOR EACH BRANCH NO. IN TABLE 8, GO TO TABLE 3 AND WRITE THE NODES THROUGH WHICH IT PASSES INTO TABLE 9
WAS 1ST BRANCH NO. Z IN TABLE 8 NEGATIVE ?
NO
YES
DELETE 2ND NODE NO. IN TABLE 9 AND ALL REPETITIONS OF NUMBERS
9
DELETE 1ST NODE NO. IN TABLE 9 AND ALL REPETITIONS OF NUMBERS
9
9
FOR EACH NODE NO. LEFT IN TABLE 9 GO TO TABLE 6 UNDER NODE COLUMN AND WRITE DOWN THE LINKS WHICH PASS THROUGH THE NODE.
IF LINK NO. IS (±P) ADD A (±1) TO THE PTH ROW OF THE ZTH COLUMN OF T MATRIX.
ARE ALL COLUMNS OF T MATRIX COMPLETED ?
YES
NO
END OF PROBLEM
PUT Z + 1 INTO Z. SET i = 1
6

Fig. 6. Flow diagrams for selecting limbs and the formation of T matrix by columns

As another example, let us obtain column 10 of our *T* matrix, that is we will define branch 10 in terms of links. From Table VII we see that −10 appears by itself so our limb is defined as just −10. This corresponds to Table VIII for our previous case. From Table III we obtain nodes 1 6, this corresponds to Table IX for the foregoing case. The number 1 is deleted and we have just the 6 to look up in Table VI. Through node 6 we have −link 3 and −link 4; this gives column 10 as listed in Table X.

Let us suppose that we wish to obtain column 14 of the *T* matrix. The second line of the right hand column of Table VII begins with 14 so that our limb branches are 14 and 17 since 17 appears by itself in line 7 of Table VII. Using Table III, our limb nodes are 3, 10, 3, 8. Again, since node 10 is the first node in the limb composed of branches 14 and 17, it is not used in defining the current in branch 14. As before, this can be verified by writing the appropriate current equations from Fig. 3. Thus, the second number, 10 in this case, is deleted from our limb nodes. Further we delete the second 3 from our limb nodes which leaves us with nodes 3 and 8. We now proceed to Table VI and see that there are no links through node 3 and only one through node 8. It is −8. $A+1$ is now placed in row 8 of column 14. It should be pointed out that writing column 14 is equivalent to solving the following equation for $B14$:

$$B14+L8=0$$

and the reason for the sign change is thus apparent.

The rule is this: Having obtained the limb nodes, we have two cases:

1. If the first branch number of the limb branch set is negative, the first node is deleted from the limb node set as well as all repetitions appearing in the limb node set. Plus 1's and minus 1's are added to the proper row of our branch column of the *T* matrix just as the links appear as plus or minus in Table VI.
2. If the first branch number of the limb branch set is positive, we delete the second node from the limb node set as well as our repetitions. This time, we pick out all links from Table VI and change their signs before adding them to the proper row of the *T* matrix.

It is possible that having chosen the set of limb nodes, we may proceed to Table VI and find that two different nodes pick out the same link. In this case the link will appear with opposite signs and the +1 and −1 when added to the proper place in our column will have the effect of putting in a zero at this location. Our *T* matrix can never contain a number different than 0, 1, or −1.

The complete *T* matrix for our sample network is shown in Table XI, while the block diagrams showing the logic for obtaining Table VI and VII appear as Fig. 5, and that for obtaining the limb nodes from Tables VI and VII is shown in Fig. 6.

FORMATION OF A TRANSFORMATION MATRIX T BY ROWS

The attempt to find a logic for formation of a transformation matrix by rows in many ways seems more natural than

Table XII

Node	Connected to Node	By Means of Branch
1	4	12**
1	6	1
1	9	13**
1	9	16*
1	10	14
2	4	10
2	5	5
2	6	2**
2	6	17*
2	7	3
2	10	9
3	8	7
3	10	8
4	1	−12**
4	2	−10
4	9	11*
5	2	− 5
5	7	4
5	8	6
6	1	− 1
6	2	− 2**
6	2	−17*
7	2	− 3
7	5	− 4
7	10	15
8	3	− 7
8	5	− 6
9	1	−13**
9	1	−16*
9	4	−11*
10	1	−14
10	2	− 9
10	3	− 8
10	7	−15

** First element in forming loop.

* All other elements in forming loop.

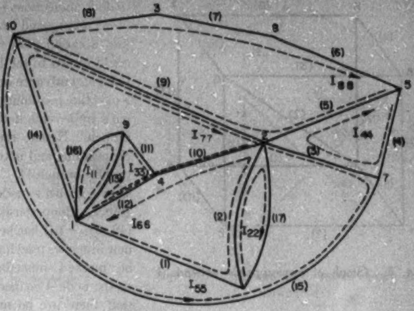


Fig. 7. Loop diagram for network 2

Table XIII

Loop No.	Branch No. 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−1	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1
3	0	0	0	0	0	0	0	0	0	0	1	1	−1	0	0	0	0
4	0	0	−1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	−1	0	0	0	0	0	1	0	0	0	0	0	−1	0	0
6	1	−1	0	0	0	0	0	0	0	1	0	−1	0	0	0	0	0
7	0	0	0	0	0	0	0	0	−1	1	0	−1	0	1	0	0	0
8	0	0	0	0	−1	−1	+1	−1	1	0	0	0	0	0	0	0	0

attempting the method described in the preceding section. The authors have studied this approach and present the results of this study in the following paragraphs. There do appear to be certain inherent differences in the matrices obtained by the two methods. A discussion of these differences is perhaps best reserved until after a description of the method is presented.

This logic will be domonstrated for the network graph of Fig. 2. The data required to form the matrix are contained in Table I. The branches of the network are not separated into links and tree branches in the procedure to be described here. The following ideas will guide the formation of the matrix by rows:

1. The number of rows (independent loop equations) required is $[b-(n_t-1),]$ in this case $[17-(10-1)]=8$.

2. Each branch in the network must appear somewhere in one of the loop equations, meaning that there must be at least one nonzero term in each column of the T matrix.

3. The matrix will be formed using the least number of nonzero elements in each row. That is, all the loop equations involving just two branches will be formed, then all those involving three branches, and so forth. The process of finding new rows will be stopped when enough rows of the matrix have been formed to meet the requirement stated in item 1 of the foregoing. The procedure will be such that the requirement stated in item 2 of the foregoing will automatically be satisfied when the proper number of rows has been written.

To begin, the information of Table I is used to form Table XII. The blocks of data in Table XII simply show for each node of the system the manner in which that node is connected to other nodes. For instance node 1 is connected to node 4 by branch 12, to node 6 by branch 1, to node 9 by branches 13 and 16, to node 10 by branch 14. It has been stated earlier that the reference positive direction for current, voltage drop, or voltage source will be assumed to be from the low numbered node to the high numbered node for every branch. This accounts for the negative signs which are associated with some branches in Table XII. Note for instance that node 7 is connected to node 2 by branch (−3), to node 5 by branch (−4), and to node 10 by branch 15. The necessity for employing these negative signs in this manner will soon be apparent. Every branch appears twice in Table XII, and while it is not absolutely essential to rely on having the data in the form of this table, some clarity is believed to result.

With Table XII available, attention is turned to writing all the rows which contain just two nonzero elements. These loops containing two branches occurs only if parallel branches exist. To select the rows with two nonzero elements, proceed as follows. Begin at the top of Table XII and note that node 1 is connected to node 4 by means of branch 12. Then transfer to the node 4 connections to determine whether or not node 4 is connected to node 1 by any branch other than (−12). Obviously, the branches (12, −12) do not close a loop, and a programmer will need to provide a simple check to avoid this absurdity. Node 4 is not otherwise connected to node 1, so return to the node 1 connections and progress downward one line in Table XII and find node 1 connected to node 6 by branch 1. Transfer to the node 6 connections and determine that node 6 is not otherwise connected to node 1. Return to the node 1 connections and find that node 1 is connected to node 9 by branch 13. Transfer to the node 9 connections and determine that 9 is connected to 1 by branch (−16). The first loop containing two branches then consists of lines (13, −16).

At this point, we mark the first element used in forming this loop by a double asterisk and all other elements used in forming the loop by a single asterisk. The purpose in so marking these branches is to be able to make the selection of loops

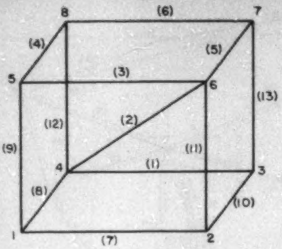


Fig. 8. Graph of nonmappable network 3

as the process continues without the necessity for checking to be sure that the loops chosen are independent. The purpose of the two different markings will be explained shortly. For the present, we will simply state that new rows will be begun using unmarked elements. Marked elements may be used elsewhere in forming a loop, but not as the first element. The second use of an already marked element does not alter the mark previously given that element.

Return to the node 1 connections and sensing the mark associated with branch 16, skip to the next line and proceed as outlined. By this procedure, the one remaining row with two nonzero elements will be found to correspond to the loop formed by the branches (2, −17).

When the table has been completely searched for two element rows, the search for rows containing three nonzero elements begins. Again starting at the top of the table, trace from 1 to 4 by the way of branch 12, then from 4 to 2 by the way of branch (−10), and then search the node 2 connections for a return to node 1. None exist, and so a new attempt to find a closed path must be made. One could return to the list of node 1 connections and advance one line, but this overlooks the possibility of finding a closed path by leaving node 4 using a different branch. Hence, return to the node 4 connections and note that 4 is connected to 9 by branch 11. The node 9 data will show a return to node 1 by branch (−13). Hence, a loop consisting of the branches (12, 11, −13) has been discovered. The new branches used for the first time should be marked immediately. Return now to the node 4 connections and determine that there are no more branches out of node 4 to be considered at this time.

Then return to the node 1 data and advance one line. By repeating the procedure just outlined, one determines that a row containing three nonzero elements cannot be obtained starting with branch 1. No attempt to start new rows with branches 13 or 16 will be made since thay are marked. The remaining three element rows are (5, 4, −3) and (9, −15, −3).

Of course the number of rows written is tabulated, and at the end of the writing the rows with three nonzero elements, there are still three rows to be written, so a search is begun for the rows containing just four nonzero elements. Following the procedure outlined in the foregoing. one discovers the loops corresponding to (1, −2, 10, −12) and (14, −9, 10, −12). The remaining row contains 5 nonzero terms and corresponds to the loop (7, −6, −5, 9, −8). The T matrix for this network using this procedure is shown in Table XIII.

The reason for the double asterisk is not brought out in this example. This is due to the fact that we were able to begin the eighth row with an element that had not previously been used. The decision as to where to begin the eighth row in order to obtain an independent loop was straight forward. However, suppose that the seventh row had included the last remaining unused element. Then some decision as to the point of beginning the eighth row would be required. By eliminating all the lines in Table XII which have been marked with a double asterisk, we would be able to begin the process all over with no concern that the rows yet to be written might not be independent.

There are some differences between this T matrix and the one written by columns. Note that there is no renumbering of branches. The loop currents, shown in Fig. 7, are not each identical to one of the branch currents. There is no 8 by 8 identity part of this matrix. It does, however, have the important properties associated with the matrix written by columns, namely

$$TV_B - TE_B = 0$$

and

$$T'I_L = I_B$$

where V_B and E_B are branch voltage drop and branch voltage source column vectors, and I_D and I_B are loop current and branch current column vectors.

APPLICATION TO A NETWORK NOT MAPABLE ON A SPHERE OR PLANE

The logic described in the two preceding sections is not restricted in its application to networks mapable on a plane or sphere. The cube with a diagonal shown in Fig. 8 may be used to demonstrate that no difficulty is encountered in forming a transformation matrix by either method. The matrix resulting from formation by columns is presented in Table XIV and the one by rows is presented in Table XV.

Table XIV

Link No.	Branch No. 1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	1	−1	0	1	0	0	0
2	0	1	0	0	0	0	−1	1	0	0	−1	0	0
3	0	0	1	0	0	0	−1	0	1	0	−1	0	0
4	0	0	0	1	0	0	0	−1	1	0	0	−1	0
5	0	0	0	0	1	0	0	0	0	−1	1	0	−1
6	0	0	0	0	0	1	1	−1	0	1	0	−1	1

Table XV

Loop No.	Branch No. 1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	1	−1	0	1	0	0	0
2	0	0	1	0	0	0	−1	0	1	0	−1	0	0
3	0	0	0	0	−1	0	0	0	0	1	−1	0	1
4	0	1	0	0	0	0	−1	1	0	0	−1	0	0
5	0	0	0	−1	0	0	0	1	−1	0	0	1	0
6	−1	0	0	0	0	1	0	0	0	0	0	−1	1

Conclusions

Methods have been described for forming a transformation matrix T by columns and by rows. Such a matrix contains the information regarding network configuration which is needed to establish the required number of independent mesh equations. The methods as presented are suitable for programming a digital computer. Experience in using T matrices formed by the two methods may very well establish a preference for one method over the other for certain types of problems. Considering several imme-

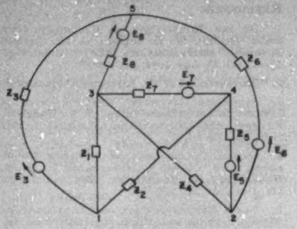


Fig. 9. Network 4

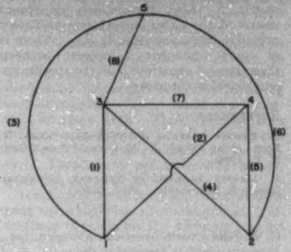


Fig. 10. Graph for network 4

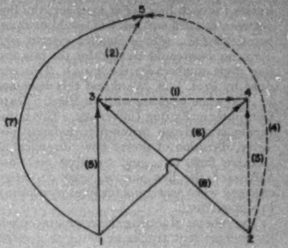


Fig. 11. Link and tree branch graph for network 4

diate applications, the authors chose to work with the formation of the matrix by columns, and a program for obtaining the matrix has been written for and used with the IBM *704* located in the Analytical Department of Westinghouse in East Pittsburgh. The following times were noted:

No. of Nodes	No. of Branches	Machine Time in Seconds
37	57	27
65	92	35
95	151	50

The times recorded here include reading data from cards and writing the final matrix on tape. The large portion of the time was used in reading cards and writing the answer tape.

Appendix

The purpose of this appendix is to show a complete solution of a network problem using a transformation matrix T to establish the network relations. The network solved here is shown in Fig. 9.

Suppose the network constants and the impressed voltages in per unit are as in Table XVI. The graph of the network will be as shown in Fig. 10, and Table I for this network will be as follows:

Line Nos.	Branch Node Designations	
1	1	3
2	1	4
3	1	5
4	2	3
5	2	4
6	2	5
7	3	4
8	3	5

The network will have the tree branches and links shown in Fig. 11. The new branch numbers are included with the table of constants. Using the method of columns we obtain the following T matrix.

Link Number	Branch Number 1	2	3	4	5	6	7	8
1	1	0	0	0	1	−1	0	0
2	0	1	0	0	1	0	−1	0
3	0	0	1	0	1	−1	0	−1
4	0	0	0	1	1	0	−1	−1

The Z matrix for this network will be as follows:

$$\begin{Vmatrix}
1+j0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.2+j0.3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.2-j0.5 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.8+j1.0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.1+j0.3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05+j0.15 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.2+j0.6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1+j0.5
\end{Vmatrix}$$

Following the procedure outlined in the text, the equilibrium equations are:

$$TE_B = TZ_B T' I_L$$

$$\begin{Vmatrix}
-1.75+j0 \\
-1.7-j0.18 \\
-0.7-j0.1 \\
-1.1+j2.62
\end{Vmatrix}
=
\begin{Vmatrix}
1.15+j0.65 & 0.1+j0.3 & 0.15+j0.45 & 0.1+j0.3 \\
0.1+j0.3 & 0.5+j1.2 & 0.1+j0.3 & 0.3+j0.9 \\
0.15+j0.45 & 0.1+j0.3 & 0.45+j0.45 & 0.2+j0.8 \\
0.1+j0.3 & 0.3+j0.9 & 0.2+j0.8 & 1.2+j2.4
\end{Vmatrix}
\begin{Vmatrix}
I_1 \\
I_2 \\
I_3 \\
I_4
\end{Vmatrix}$$

These equations have the following solutions:

$I_1 = -1.23633+j0.666294$
$I_2 = -1.167166+j0.537270$
$I_3 = 0.151991-j1.184075$
$I_4 = 1.156179+j0.961608$

The remaining branch currents may be determined using rows 5, 6, 7, and 8 of T' and the link current vector. The values obtained are:

$I_5 = -1.09533+j0.98109$
$I_6 = +1.08434+j0.51778$
$I_7 = 0.01098+j1.49888$
$I_8 = -1.30817+j0.22246$

Table XVI

Original Branch Numbers	Branch Impedances, Pu	Impressed Voltage, Pu	New Branch No.
1	0.1+j0.3		5
2	0.05+j0.15		6
3	0.2+j0.6	1.0+j0.08	7
4	0.1+j0.5		8
5	0.2−j0.5	−0.7−j0.1	3
6	0.8+j1.0	−0.1+j2.70	4
7	0.1+j0.2	−1.75+j0	1
8	0.2+j0.3	−0.7−j0.1	2

Reference

1. INTRODUCTORY CIRCUIT THEORY (book), Ernst A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1953, pp. 1–58.

◆

Discussion

E. E. George (Ebasco Services Inc., Little Rock, Ark.): The systems of line and bus numbering used by the authors provides completely independing numbering for both groups. There is an improvement in the numbering system that could be made to simplify the explanation as well as the computer coding. In this new procedure busses could be assigned three digit numbers and then the lines could be denoted by a number with twice as many digits, comprised of the numbers of the two terminal busses for each line. For instance, the line between bus 036 and bus 243 would be numbered 036243. (On a smaller system two digit numbers could be used for the bus and four digit numbers for the lines.) This was done by Prof. J. B. Ward and this discusser in 1956 when working with an Electro-Data Computer at Purdue University. This numbering system can be used to indicate polarity, direction of power flow, direction of reactive flow, etc., and the first and last halves of the line number can be interchanged when searching for lines connecting to a given bus.

The scheme of letting the computer select the independent tracks or loops in a mesh system is a valuable development and the authors are to be highly praised. This task has always been the most tedious operation in working with a mesh network, and it has been a major source of errors.

The selection of independent, necessary, and sufficient equations in a large mesh network is so difficult by manual methods that the mesh method has come near being discarded as a practical tool. Now that the computer can be used to do this difficult chore, we may find more uses for mesh network methods.

This discusser developed a process chart in 1937 for tracing mesh tracks on a computer, but it was never coded or applied to a particular type of computer. He, therefore, has an appreciation of the practicability and utility of such an application of Messrs. Byerly, Long, and King have described.

The authors use the word "link" to denote a branch that is not a tree branch. This discussor prefers the term "air branch" instead of "link." A mesh would then be composed of "tree branches" and "air branches," with each term being descriptive and therefore easier to remember.

L. K. Kirchmayer and **A. J. Wood** (General Electric Company, Schenectady, N.Y.): The authors' presentation of a logical technique for the computer generation of a connection matrix is interesting and will undoubtedly be of value to engineers and others concerned with the digital computer solution of electric networks.

The equations under the section on "Review of Some Geometrical Aspects of Networks" of the paper are in essence a specific application of G. Kron's general laws of transformation.[1,2] These laws of transformation may be stated as:

if

$$\bar{I}^{\text{old}} = \bar{C}\bar{I}^{\text{new}}$$

then

$$\bar{E}_{\text{new}} = \bar{C}_t{}^*\bar{E}_{\text{old}}$$

$$\bar{Z}_{\text{new}} = \bar{C}_t{}^*\bar{Z}_{\text{old}}\bar{C}$$

where the subscripts old and new refer to the old and new circuits respectively and I, E, and Z denote the corresponding current, voltage, and impedance matrices. The matrix $\bar{C}$ is often referred to as Kron's connection matrix. The matrix $\bar{C}_t{}^*$ is the conjugated transpose of $\bar{C}$.

If the branch currents and voltages are designated by the matrices $\bar{I}^{\text{old}}$ and $\bar{E}_{\text{old}}$ and the loop or link quantities by $\bar{I}^{\text{new}}$ and $\bar{E}_{\text{new}}$, then the authors' specific transformation matrix T' corresponds to Kron's connection matrix $\bar{C}$. Also, the branch impedance matrix of the paper corresponds to Kron's primitive impedance matrix $\bar{Z}_{\text{old}}$. There is of course complete agreement between these general equations and the specific example used in the paper.

REFERENCES

1. TENSOR ANALYSIS OF NETWORKS (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1939.

2. MATRIX ANALYSIS OF ELECTRIC NETWORKS (book), P. LeCorbeiller. *Ibid.*, 1950.

G. W. Bills (North American Aviation, Inc., Downey, Calif.): Inevitably the marriage of topological methods and digital computers for the automatic solution of electric networks had to occur. In 1938 the discusser was a graduate student under Professor Cramlet at the University of Washington when the latter began his research into the geometrical properties of electric networks and which resulted in the classic paper by Prof. Cramlet and Ingram.[1] Just previous to this time Gabriel Kron began publishing his papers on tensor analysis of networks.[2] Unfortunately there were no high-speed, large-memory electronic digital computers available then to make it practical to apply such methods to large electric networks. The papers by Wong[3] and Glimn[4] were among the first to do this. The present paper is of interest because it supplies a needed tool when large electric networks must be analyzed by the mesh approach; and it demonstrates the effectiveness of the topological technique when combined with a computer such as the IBM *704*.

This paper rightfully recognizes that iterative methods are just as applicable to the solution of mesh current equations as to nodal equations. Strangely, most papers which apply mesh analysis to power system analysis have resorted to matrix inversion, a process usually inferior to iterative techniques for large systems.

I should like to point out that mutual inductances can readily be included in nodal equations, as demonstrated in a paper[5] presented at the AIEE Pacific General Meeting in August 1958. For this reason it is not necessary to resort to mesh analysis when mutual inductances are present in a system.

A little over a year ago our Analysis Unit began studying Guilleman's excellent study[6] for the purpose of applying topology to the automatic coding of networks with an IBM *704*. Unfortunately the cancellation of the NAVAHO missile program brought a sudden and prolonged halt to our digital programing activities. For this reason we are especially grateful to know that a digital program such as the one described in the paper is in existence. Is this program now listed with SHARE or at least available to IBM *704* users?

Any systems engineer or circuit theory analyst should appreciate the simplicity of this program for obtaining a T matrix. Because the method is so clear, it appears that clerical help could be quickly taught to number nodes and branches, form Table I, and then place the values on data sheets for entry into the digital computer, thus releasing engineer for more creative tasks.

Have any networks been found where the program could not form a correct matrix?

REFERENCES

1. ON THE FOUNDATIONS OF ELECTRICAL NETWORK THEORY, W. H. Ingram, C. M. Cramlet. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 23, Aug. 1944, pp. 134–55.

2. See reference 1 of Mr. Kirchmayer's discussion.

3. AUTOMATIC NETWORK ANALYSIS WITH A DIGITAL COMPUTATION SYSTEM, S. Y. Wong, M. Kochen. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 172–76.

4. DIGITAL CALCULATION OF NETWORK IMPEDANCES, A. F. Glimn, R. Habermann, Jr., J. M. Henderson, L. K. Kirchmayer. *Ibid.*, vol. 74, pt. III, Dec. 1955, pp. 1285–97.

5. NODAL REPRESENTATION OF LARGE COMPLEX ELEMENT NETWORKS INCLUDING MUTUAL REACTANCES, J. C. Siegel, G. W. Bills. *Ibid.*, vol. 77, pt. III, 1958 (*Paper no. 58-1023*).

6. INTRODUCTORY CIRCUIT THEORY (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1953.

Ernest J. Schubert (Burroughs Corporation, Paoli, Pa.): This paper describes a computer analysis for networks of higher complexity. Branches are somewhat confusingly denoted in four different ways by line no. (Table I), by both the nodes, link and tree numbers equivalent to a "new branch no." (Table V). No consistency can be seen in numbering nodes or branches.

Better efficiency in dealing with similar problems may be gained by organizing the concept differently:

1. Only one branch shall be considered connecting any two nodes. Parallel paths between the same nodes have to be combined in a single equivalent branch.

2. The number of nodes N determines the possible complexity of a net, because the number of branches B cannot exceed the combinations of n things taken two at a time.

$$B \leq N(N-1)/2$$

3. In opposition to the section on "Application to a Network not mapable on a Sphere or Plane" it should be noted that any net may be mapped in a plane. Intersecting branches might occur for more than three nodes without being relevant for the solution. If drawing a netgraph at all, nodes shall be arranged approximately in a circle, numbering the nodes in sequence.

4. Nets comprising N nodes might be formed depending on the number of branches B between two extremes:

complete web $W=(N-1)N/2$
tree $T=N-1$

Branches in excess of those forming the tree are links L. Their number has an upper bound.

$$L = B - T \leq W - T$$

5. A net comprising N nodes is completely described by a matrix with N rows i and columns j describingo rigin and destination of any possible branch by a branch impedance $Z_{ij} = Z_{ji}$ considering $i<j$ only. Because of symmetry and insignificance of Z_{ii} (node connected to itself) $N(N$-$1)/2$ storage locations in the computer are required only.

6. To determine without tabulation links and loops by computer routine an image matrix S is formed with:

$S_{ij}=0$ for $Z_{ij}=0$
$S_{ij}=1$ for $Z_{ij}\neq 0$

Forming S^2, S^3 gives the required information for computer instruction as follows:

6.1...$S_{ii}^2=C$...the node i is connected to C other nodes
6.2...$S_{ii}^3=0$...the node i is single ended and does not belong to any loop
6.3...$S_{ii}^3=2$...the node i is part of one loop
6.4...$S_{ii}^3=4$...the node i is part of two loops, etc.

From S_{ii}^3 and S_i a subroutine may directly eliminate links to form a tree.

It is regrettable that the pioneering efforts of G. Kron in his classic book[1] have not been mentioned as the most pertinent reference.

REFERENCE

1. See reference 1 of Mr. Kirchmayer's discussion.

Martin J. Lantz (Bonneville Power Administration, Portland, Oreg.): The authors are to be commended for this fine paper. This method is applicable to short-circuit solution of power system networks and does include mutual coupling between circuits for zero-sequence circuit problems. The discusser has had experience in this type of problem for the IBM *650* digital solution.[1] Since the reference 1 paper was published the average fault solution time has been reduced by a factor of three and considerable simplification made in the problem preparation.

It is my feeling that the mesh approach to this problem is the more practical for the average utility engineer.

In prior discussion with the author, the then assumed method of establishing faults was to include an element for every fault position which practically doubles the size of the impedance matrix. This large matrix must be inverted once and the large inversion multiplied by the T matrices for each fault. It seems that the subsequent matrix multiplications with the very large matrix inversion may result in greater solution times than the method described in reference 1 which requires an inversion of a matrix of half the size for less than one third of the faults in the revised program.

It is not clear how the authors intend to represent other than bus faults, such as faults on either side of a line breaker after the breaker has opened for circuits involving mutual coupling. These faults are important for relay work. Three terminal lines require a sequential opening of circuit breakers which further complicates the solution. These faults are all readily accomplished in the reference solution.

It is hoped the authors will continue their program to cover an automatic solution of all positions of faults for the mutually coupled networks.

REFERENCE

1. DIGITAL SHORT-CIRCUIT SOLUTION OF POWER SYSTEM NETWORKS INCLUDING MUTUAL IMPEDANCE, Martin J. Lantz. *AIEE Transactions*, vol. 76, pt. III, 1957 (Feb. 1958 section), pp. 1230–35.

R. T. Byerly, R. W. Long, and C. W. King: The authors appreciate the many fine comments presented in the discussions.

Mr. George is to be thanked for clearly indicating the need for an automatic method of obtaining the transformation matrix T and hence has reemphasized the main purpose of the paper. His comments with regard to the numbering of the lines and nodes of a network is well taken and might be used to advantage by persons interested in trying to improve the method for developing a T matrix.

Dr. Kirchmayer and Mr. Wood have nicely shown the generalization of the T matrix to the complex field. They have further pointed out that the T matrix is the same as the connection matrix C as used in the work of Kron and Le Corbeiller.

The remarks of Mr. Bills are timely and the relation of his experiences has helped to fill in the background of the problem. The authors feel that certain problems may be better handled using the nodal analysis approach so that it is well to note that mutual inductances may be included in this analysis.

Mr. Schubert has indicated a method for obtaining the tree branches and links for an electric network. This could be used to solve the first part of the problem in a different manner. It can be expected that since there are many people who will become more interested in the T matrix and its application that even better methods will be forthcoming. The answer as to what is best will not be known until the ease of preparing data has been taken into account and the machine times for these methods are made available.

It has been the desire of the authors to keep the labeling of the network branches and nodes as flexible as possible. It seems that to require a program user to even parallel branches tends away from this desire. It is true that the computer is forced to do a relabeling in the authors' method but this is rapidly accomplished. Further, when any answers are printed such as for a short-circuit solution, the results are printed using the original labels assigned by the user.

In a previous discussion with Mr. Lantz the authors indicated some of the steps they were following in applying the T matrix to the solution of the short-circuit problem. Since that time a solution to this problem has been completed and the method has been described in a paper[1] to be presented at the Power Industry Computer Application Conference. It is hoped that this paper will answer the questions raised by Mr. Lantz.

In further answer to Mr. Bills, it should be stated that the *704* program using the method described in the paper is available for use by the electric utility customers on Analytical Department facilities in East Pittsburgh. The use of this program has been quite successful and has not failed to form the correct T matrix. In fact the logic of the program though it might seem complicated, is very straightforward. The reasoning is so clearcut that the method is applicable to all types of networks, even those not topologically mappable on a plane or sphere. An example of this type of network is shown in Fig. 8 of this paper.

REFERENCE

1. DIGITAL CALCULATION OF POWER SYSTEM NETWORKS UNDER FAULTED CONDITIONS, R. T. Byerly, C. J. Baldwin, Jr., R. W. Long, C. W. King. *AIEE Transactions*, vol. 77, pt. III, 1958 (*Paper no. 58-1099*).

Magnetic-Drum Components for High Storage Density

KYONGHI HONG
NONMEMBER AIEE

THE INFLUENCE of coercivity H_c and retentivity B_r of the recording medium has been studied by numerous authors in the field of magnetic recording. In 1946 Camras[1] considered the demagnetizing field of a bar magnet comprising the record medium and concluded that the reproduced voltage was proportional to B_r for long wavelengths and to H_c for short wavelengths. In 1947 Kornei[2] observed that a medium of high values of H_c/B_r gave better high-frequency response, while Holmes[3] concluded that in sound recording the ratio H_c/B_r should be greater than 0.1 and less than 1. Wetzel[4] also expressed a view similar to that of Camras on H_c and B_r and presented data supporting it in 1948. Begun[5] pointed out in 1949 that H_c was the most important factor in magnetic recording although he stated in 1955[6] that high H_c does not add materially to the medium's capability of retaining a short-wavelength signal unless B_r can be increased at the same time. In 1953 Daniel[7] reached the conclusion that as the wavelength approaches zero the playback voltage becomes proportional to $1/(1+B_r/H_c)$, while McGuigan[8] pointed out that the area recorded on the medium at the

---

Paper 58-929, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 10, 1958; made available for printing May 7, 1958.

KYONGHI HONG resides in Cambridge, Mass.

time of writing spreads out over an area larger than the area directly under the pole tip of the head, and that the extent of this spreading depends on H_c of the medium, without specifying how.

The resolving power of the playback head has been discussed by Begun;[8] Fuller, Husman, and Kelner;[9] and Hoagland.[10] Begun mentioned only the gap width as a factor determining the resolving power of the head and pointed out that the effective gap length is somewhat larger than the actual gap length, usually by about 50 to 100% for short gaps and down to 10% for long gaps. Fuller, Husman, and Kelner regarded the spacing between the head and medium as well as the gap length as factors determining the resolving power. Hoagland, on the other hand, believed the thickness of the medium as well as gap length and spacing to be the factors determining the resolving power.

Comprehensive formulas for maximum storage density of the medium, the playback voltage, and the resolving power of the head are derived in this paper to augment, correct, and resolve the conflicting viewpoints. The maximum storage density is the reciprocal of the minimum wavelength that the medium is capable of retaining. The wavelength is the velocity of the medium divided by the frequency of the writing current. Although these formulas should be applicable to any magnetic recording, audio or video, the main concern of this paper is the magnetic components of a magnetic drum where information is handled in binary digits (bits) and requires a distinction between only two magnetic states, erasing and biasing being unnecessary.

Maximum Storage Density

It is shown in the literature[11] that the second quadrant of a hysteresis loop can be expressed as

$$B/B_r = \frac{1+H/H_c}{1+(H/H_c)(B_r/B_s)} \tag{1}$$

where B is the induction, B_r the retentivity, H the magnetizing force, H_c the coercivity, and B_s the saturation induction as shown in Fig. 1. It is seen in this equation that B becomes B_r when H is zero, and H becomes $-H_c$ when B is zero. Westmijze[12] has shown that, in tape recording, the demagnetizing field of a longitudinally magnetized bar magnet comprising the tape can be expressed as

$$-H/B = -1/\mu = L/4\lambda$$

or

$$-H = BL/4\lambda \tag{2A}$$

where λ is the wavelength and L is the tape thickness magnetized uniformly throughout, while for perpendicular magnetization

$$-H/B = -1/\mu = (1-\pi L/\lambda)/4\pi$$

or

$$-H = B(\lambda-\pi L)/4\pi\lambda \tag{2B}$$

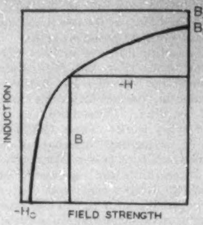


Fig. 1. Demagnetization curve

Combining equations 1 and 2(A) gives, after some rearrangement,

$$x = y(hy-1)/(y-1) \tag{3}$$

$$y = B/B_r$$

$$x = 4\lambda H_c/LB_r$$

$$h = B_r/B_s$$

x is plotted against y for various values of h in Fig. 2. L in Westmijze's derivation was the thickness of a uniformly magnetized tape. If the thickness of the tape be infinite from a practical point of view, then L may be regarded as the depth of magnetic penetration. In Fig. 2 it is seen that the linear relation between x and y holds when x is less than 0.2, or when h is equal to 1. When x is less than 0.2 the wavelength is less than half a mil, 1.27×10^{-3} cm (centimeter). When the wavelength is greater than half a mil and B_r is not equal to B_s the linearity fails and B depends on B_r. For the short-wavelength region which is the primary concern of this paper, the induction is

$$B = 4\lambda H_c/L \text{ or } 4\pi\lambda H_c/(\lambda-\pi L) \tag{4}$$

The relation between λ and L can be estimated as follows: if the tape is magnetized with both perpendicular and longitudinal components as is the case in practice, then the tape may be regarded as a chain of flux-closure domains as shown in Fig. 3. It is shown in the literature[13] that the relation between λ and L in Fig. 3 is

$$(\lambda/2)^2 = 2E_wL/K \tag{5}$$

where K is the anisotropy constant and E_w is the Bloch wall energy per-unit area, which is about 1 to 5 ergs/cm² for 180-degree-spin orientation. Equation 5 states that the magnetic penetration into the medium is proportional to the square of the wavelength.

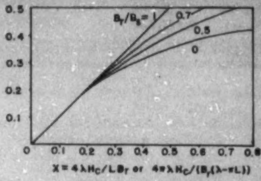


Fig. 2. Calculated curve of B/B_r versus $4\lambda H_c/\pi LB_r$ or $4\pi\lambda H_c/B_r(\lambda-\pi L)$ for various values of B_r/B_s

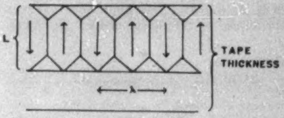


Fig. 3. Flux-closure domains of the medium

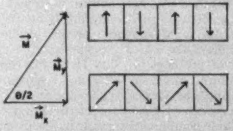


Fig. 4. Magnetization of medium with different angles θ and π between domains

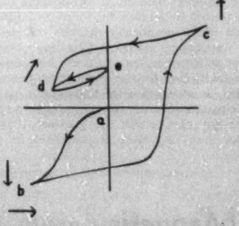


Fig. 5. Magnetic state of a point on a medium varied by recording and reproducing heads. Vectors illustrate the direction of magnetization

The flux-closure model in Fig. 3 gives some insight into the relation between λ and L, but it cannot be used for the study of the phenomena at the surface of the medium, for it has no poles. To cal-

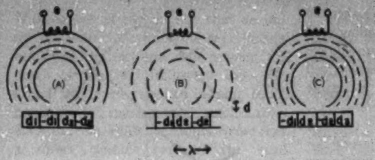


Fig. 6. Reading-head position for maximum (solid line) and minimum (dashed line) flux density. The mean circumference of the head is assumed constant

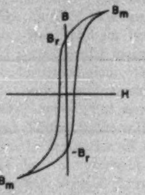


Fig. 7. Magnetic state of playback head varied by its position relative to record medium

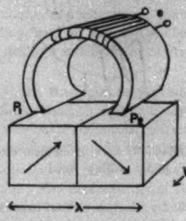


Fig. 8. Head-medium geometry

culate the maximum storage density of the medium, the tape is assumed to be magnetized as in Fig. 4. The energy content of the medium F_v, per-unit length of volume V, is

$$F_v = VB^2/8\pi\mu = (VL/32\pi\lambda)(4\lambda H_c/L)^2$$

$$= \lambda VH_c^2/2\pi L \qquad (6)$$

Since the magnetization of the magnetically hard recording medium takes place by the rotation of domains, angle θ in Fig. 4 is a measure of the magnetization, i.e., $\theta/\pi = B/B_s$, and the Bloch wall energy E_w' is proportional to the total change in spin direction θ. Therefore the sum of wall energies within unit length of the medium of cross-sectional area A with U bpc (bits per cm) would be

$$F_w = 2AUE_w' = 2AUE_w\theta/\pi = 2AUE_wB/B_s$$

$$= 2AUE_w(4\lambda H_c/LB_s) \qquad (7)$$

Increase in the storage density U can continue until F_w becomes equal to F_v. Such an equality gives the following formula for the maximum storage density

$$U = H_cB_s/16\pi E_w \qquad (8)$$

It can be readily shown that equation 8 is true for perpendicular magnetization also if $4\pi\lambda H_c/[B_r(\lambda - \pi L)] < 0.2$ or $B_r/B_s = 1$. If $H_c = 500$ oersteds, $B_s = 6{,}000$ gauss, $E_w = 2$ ergs/cm², then $U = 30{,}000$ bpc, and the minimum wavelength the medium is capable of retaining is $\lambda_c = 1/U = 0.3333 \times 10^{-4}$ cm.

Playback Voltage

A formula for playback voltage is of practical value in understanding the operation of magnetic recording. The medium is assumed to be demagnetized initially and is at the magnetic state, of **a** in Fig. 5. As a point on the medium comes under a writing head, its magnetic state is changed from **a** to **b** when the point is exactly under the leading edge of the gap. If B_s and ampere-turns of the writing head be great enough and if the medium-head spacing be negligible, the magnetic state at **b** is that of perpendicular saturation. The point is longitudinally saturated when it comes halfway between the gap edges, and varies from **b** to **c** when it comes under the trailing edge of the gap, and from **c** to **d** after it leaves the head. The value of B at **d** is less than B_r because of a demagnetizing field, but it becomes almost doubled to **e** when it comes under a reproducing head because the reproducing head reduces the demagnetizing field.

The magnetic state of the reading head (also called playback head or reproducing head) can be understood from Figs. 6 and 7. When the record medium moves under the reading head whose effective gap length is not equal to an integral multiple of the wavelength, the magnetic state of the head will be at $-B_m$ of Fig. 7 when it is at the magnetically asymmetric position of Fig. 6(A), at $-B_r$ when at the magnetically symmetric position of Fig. 6(B), and at B_m when at the magnetically asymmetric position of Fig. 6(C). In Fig. 6, position (B) is reached when the medium moves a distance of $\lambda/4$ from position (A). Position (C) is reached when the medium moves a distance of $\lambda/4$ from position (B). If the velocity of the medium be v, then the output voltage from the reading head is

$$e = -NAdB/dt = -NA(-B_r - B_m)/(\lambda/4v)$$

$$= 6vNAB_m/\lambda \qquad (9)$$

The last step is based on the assumption that $2B_r = B_m$, which is somewhat arbitary but does not hurt the generality. N is the number of turns of the reading coil wound tightly around the core at a symmetric position with respect to the gap, and A is the cross-sectional area of the core. For RZ (return-to-zero) coding, however, $e = 4vNA(B_r - B_m)/\lambda = -2vNA$-$B_m/\lambda$ also occurs.

The value of B_m is to be estimated by use of the model in Fig. 4 magnified in Fig. 8 where the effective gap length is assumed to be $\lambda/2$. The magnetic potential at P_1 is called V_1, and is determined by the surface distribution and volume distribution of the pole-strength density, namely,

$$V_1 = \int_s \frac{\sigma ds}{r} + \int_\tau \frac{\rho d\tau}{r} \qquad (10)$$

where σ is the pole strength per-unit area on the surface and ρ is the pole strength per-unit volume of the medium. V_1 is equal and opposite to V_2 at P_2. The model shown in Fig. 4 is that of opposing longitudinal bar magnets plus attracting transverse bar magnets at the same time. In case of a transverse bar magnet of length L and cross-sectional area $w\lambda/2$ magnetized uniformly along its axis, the magnetic moment of the bar is $\sigma Lw/2$, and the magnetic moment per-unit volume is[14]

$$\rho = M = B/4\pi = \sigma \qquad (11)$$

Combining equations 4 and 11 modifies equation 10 as follows:

$$V_1 = V = -V_2$$

$$= Mw(\lambda/2)[(1/d+L)/(d+L/2)]$$

$$= Mw\lambda/2d$$

$$= \lambda^2H_cw/\pi Ld$$

B (or M) is assumed doubled by the shunting effect of the reading head on the basis of the image method used by Wallace.[16] Therefore B_m can be expressed as

$$B_m = (2\mu V/l)[R_g/(R_c + R_g)]$$

$$= (2\mu gV/l)/(l/\mu + g)$$

$$= \frac{2\mu gH_cw\lambda^2}{ld(\lambda - \pi L)(l/\mu + g)} \qquad (12)$$

In equation 12 the cross-sectional areas of the gap and the core are assumed to be equal. The factor $R_g/(R_c + R_g)$ is in-

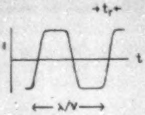


Fig. 9. NRZ current

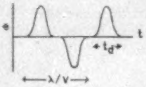


Fig. 10. Output waveform from NRZ coding

cluded because of the flux ϕ_c for the core is only part of the total flux ϕ such that $\phi_c = \phi[R_g/(R_c+R_g)]$, where R_g is the reluctance of the gap and R_c the reluctance of the core. μ is the permeability of the reading head, g the gap length, and l the magnetic path length of the core. Combining equations 9 and 12 now gives

$$e=\frac{12\lambda vNA\mu H_c w\sin(\pi g/\lambda)}{ld(\lambda-\pi L)(l/g\mu+1)}\text{ for perpendicular magnetization}\quad(13)$$

$$e=\frac{12\lambda vNA\mu H_c w\sin(\pi g/\lambda)}{ld\pi L(l/g\mu+1)}\text{ for longitudinal magnetization}\quad(14)$$

The factor $\sin(\pi g/\lambda)$ can be understood from Fig. 6(A). The gap length of the dashed line is λ and the flux through the reading head is zero because it is at a magnetically symmetrical position. A gap length 2λ is similar. In general, a gap length $n\lambda$ gives a zero field intensity, n being 0, 1, 2, 3, ..., as shown in Fig. 6(A). The maximum field intensity is obtained when $g=(n+{}^1/_2)\lambda$. This situation may be summed up by saying that the field intensity, and consequently the output voltage, varies as $\sin(\pi g/\lambda)$ and this explains the appearance of maxima and minima when the gap length is greater than one wavelength. Optimum gap length is $\lambda/2$.

In equation 13 for perpendicular magnetization the output voltage increases with the thickness L of the medium. In equation 14 for longitudinal magnetization the output voltage decreases when the tape thickness L is increased. In practice, however, the medium contains both longitudinal and perpendicular components. Therefore the output voltage is somewhere between the two expressions.

Resolving Power of the Reading Head

If the resolving power U of the medium is defined as the maximum storage density that the medium is capable of retaining as calculated in equation 8, then the the resolving power U_r of the reading head may be defined as the maximum storage density (also called packing density, cell density) that the reading head is capable of reproducing with full resolution.

It is shown in the literature[13] that the Bloch wall separating two adjacent domains oppositely saturated at 180 degrees has a thickness equal to $(kT_c/Ka)^{1/2}$, where k is the Boltzmann's constant 1.38×10^{-16} erg/degree; K the anisotropy constant which for a magnetically hard material is bH_cB_s, b being a constant close to 0.1; T_c the Curie temperature in absolute scale; and a the lattice constant which is a few times 10^{-8} cm. This thickness has to be multiplied by θ/π or B/B_s, because the record medium contains both longitudinal and perpendicular magnetization. If a numerical calculation of the ratio of this thickness to a wavelength is carried out, it turns out to be of the order of 10^{-2} for most recording mediums. Therefore the Bloch wall thickness cannot have much influence on the resolving power of the medium or the head.

A sequence of a nonreturn-to-zero (NRZ) current in Fig. 9 will have an output waveform shown in Fig. 10 as can be deduced from Fig. 6. In Fig. 10 a minimum wavelength λ_c that the reading head is capable of reproducing with full resolution satisfies the condition

$$\lambda_c/v=2t_d\quad(15)$$

The reproducing head can be looked upon as a pulse transformer whose rise time and fall time depend upon the winding inductance and the drive and load characteristics of the transformer. In Fig. 10 t_d is merely the duration time of the pulse. The primary winding is the magnetic flux from the record medium, and the magnetizing force H from a perpendicularly magnetized medium can be obtained from equation 12, namely

$$H=\frac{2gw\lambda^2H_c}{ld(\lambda-\pi L)(l/\mu+g)}\quad(16)$$

The general relation[16] between H and t_d for a constant rise time of the input pulse can be obtained from Kittel's equation of ferromagnetic resonance,[13] namely, $\omega_0=\gamma(BH)^{1/2}$, or $f=1/T=1/2t_d=\gamma H\sqrt{\mu}/2\pi$. It follows that

$$t_d=1.78\times10^{-7}/H\sqrt{\mu}\quad(17)$$

The relation between t_d and the rise time of the input pulse for a constant field intensity H can be given as a function of an optimum gap length as follows:

$$t_d=g/4v\quad(18)$$

Since neither equation 17 nor equation 18 is complete by itself, a unified expression for the duration time may be obtained by multiplying the foregoing two expressions of t_d and taking the square root:

$$t_d=2.1\times10^{-4}(g/vH\sqrt{\mu})^{1/2}$$

which, when combined with equation 15, gives

$$U_r=2{,}381(H\sqrt{\mu}/vg)^{1/2}\quad(19)$$

If $\mu=10^4$, $H=4.75\times10^{-2}$ oersted, $g=5\times10^{-5}$ cm, $v=600$ cm/second, then the resolving power of the head is $U_r=30{,}000$ bpc, and the minimum wavelength the head is capable of reading is $\lambda_c=0.3333\times10^{-4}$ cm.

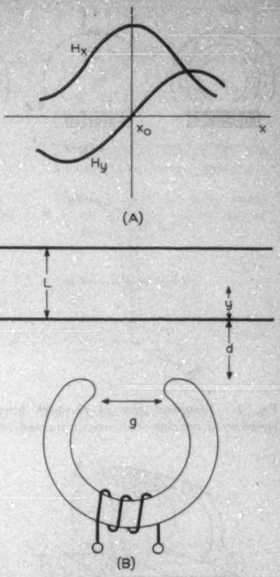


Fig. 11. Magnetic state of medium directly under writing head

Other Remarks

Formulas for the playback voltage in the short-wavelength region were previously derived by Camras,[1] Wallace,[15] Westmijze,[12] Daniel,[7,17] and Hoagland.[18] Camras believed that in the short-wavelength region the output signal increased when H_c or the reciprocal of the diameter of the recording wire was increased, although he did not explicitly show them in his equation. Daniel assumed zero spacing between the medium and the reproducing head in his derivation, but

Wallace was of the opinion that an imperfect contact between the reproducing head and the medium may account for much of the high-frequency loss. He found that even repeated lapping could not make the spacing much smaller than 5.8×10^{-4} cm, while Westmijze found that a direct contact between the head and medium was equivalent to an effective spacing of 3.6×10^{-4} cm. Westmijze's formula, like that of Wallace at unit permeability, was valid for both longitudinal and perpendicular magnetization according to his calculation. Demagnetizing fields were not considered by Wallace and Hoagland, but were studied by Camras,[1] Wetzel,[4] Westmijze,[12] and Daniel.[7] In the derivation of equations 13 and 14 for the playback voltage, the magnetic and physical characteristics of the head and medium as well as the demagnetizing fields of the medium were considered. They explicitly express the output voltage as a function of the reluctance, gap length, and spacing of the reading head as well as the coercivity of the medium independent of its retentivity.

As for the resolving power of the medium, equation 8 gives an answer to all the previous speculations on the maximum storage density of the recording medium.

Equation 19 points out that the field intensity and permeability of the reading head and the velocity of the medium contribute to the broadening effect of the output pulse, but it is also clear that the gap length of the reading head remains the most dominant factor determining the resolution. Improved gap designs have been attempted by several workers in recent years with success. Daniel and Axon,[17] in 1953 reported three different kinds of wedge-shaped gaps to give better performance in the short-wavelength region. Lufcy and Heath[19] reported in 1955 that Alfenol, whose initial permeability is 3,450 in contrast with 16,000 of Mo-permalloy, had an effective gap length about the same as its physical gap length. They obtained about 10^{-4}-cm physical gap by use of a beryllium-copper spacer. Hoagland[18] in 1956 recommended a pointed-type pole-piece design for a high storage density. Zenal[20] of the Radio Corporation of America reported in September 1956 that a video head resolving over 5,900 cycles/cm had been designed for color-television signals. In November 1956 Kornei[21] published data on a ferrite core head with a gap length of 7.9×10^{-5} cm which was spaced by silicon monoxide, and was capable of pulse-recording at densities up to 1,000 RZ pulses/cm.

The foregoing analysis of the maximum storage density leading to equation 8 was carried out from a magnetostatic point of view. If the same problem is approached from a dynamical point of view, the problem is that of diffusion. The degree of magnetic diffusion may have only a remote relation with skin depth $(\rho/\pi f\mu)^{1/2}$ where ρ is resistivity and μ permeability, but Karlqvist[22] believed that the magnetic field in the medium obeyed Fick's law of diffusion with ρ/μ as a diffusion coefficient. His graphs showed that a medium of low permeability should have less spreading (or diffusion) in longitudinal direction. The general results of his calculations, however, were similar to those published by Begun[23] in 1948. Their data are expressed in Fig. 11 and in the following equations

$$H_x=\frac{a(g+d+y)}{(d+y)[(x_0^2-x^2)^2+(g+d+y)^2x^2]}$$

$$H_y=\frac{a(x_0^2-x^2)}{(d+y)[(x_0^2-x^2)^2+(d+g+y)^2x^2]} \qquad (20)$$

where y is the depth in the medium measured from the surface of the medium, x is the longitudinal position, and a is a constant determined by the B_s and ampere-turns of the writing head. If the spacing d be negligible, the peak value of H_y may become equal to the peak value of H_x in Fig. 11(A). The situation here is similar to that of anomalous dispersion showing the index of refraction and absorption coefficient as functions of frequency,[24] or the reactance and resistance of a resonant electric circuit as functions of frequency. If the half-width of H_x curve is defined as the distance between the points at which H_x has half its maximum value, then it is approximately $(g+d+y)$ or the distance between the peaks of H_y curve. y may be equal to L, the tape thickness or the depth of magnetic penetration. For a high-density RZ coding, the velocity of the medium is selected by the relation $v/f=\lambda_c=2(g+d+L)+vt_d$. Solving this equation for v gives

$$v=2(g+d+L)/(1/f-t_d)$$

where t_d is the pulse duration time shown in Fig. 12, $1/f$ is the period of the RZ current, and λ_c is the minimum wavelength the medium or the reading head is capable of resolving. In NRZ coding, only the ratio $v/f=\lambda_c$ has to be considered.

Conclusions

1. The maximum storage density of a medium increases with its coercivity H_c and saturation induction B_s as shown in equation 8.

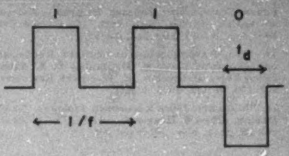


Fig. 12. RZ current

2. In a short-wavelength region, the playback voltage is proportional to the H_c of the medium and is independent of its retentivity B_r; see equations 13 and 14.

3. The resolving power of the reading head is mainly determined by the gap length, but is influenced by the field intensity and permeability of the head and the velocity of the medium also, as shown in equation 19.

References

1. Theoretical Response from a Magnetic Wire Record, **Marvin Camras.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Aug. 1946, pp. 597–602.

2. Frequency Response of Magnetic Recording, **Otto Kornei.** *Electronics*, New York, N. Y., Aug. 1947.

3. Some Factors Influencing the Choice of a Medium for Magnetic Recording, **Lynn C. Holmes.** *Journal*, Acoustical Society of America, Lancaster, Pa., May 1947.

4. Review of Present Status of Magnetic Recording, **W. W. Wetzel.** *Audio Engineering*, Mineola, N. Y., Nov. 1947; Jan. 1948.

5. Magnetic Recording (book), **S. J. Begun.** Murray Hill Books, Inc., New York, N. Y., 1949, pp. 207, 106.

6. A Survey of Magnetic Recording, **S. J. Begun.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 651–55.

7. The Influence of Some Head and Tape Constants on the Signal Recorded on a Magnetic Tape, **E. D. Daniel.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. III, 1953, p. 168.

8. Combined Reading and Writing on a Magnetic Drum, **J. H. McGuigan.** *Proceedings*, Institute of Radio Engineers, New York, N. Y. Oct. 1953, p. 1438.

9. Techniques for Increasing Storage Density of Magnetic Drum Systems, **H. W. Fuller, P. A. Husman, R. C. Kelner.** "Proceedings of the Eastern Joint Computer Conference," *AIEE Special Publication T-70*, Dec. 1954, pp. 16–21.

10. Magnetic Drum Recording of Digital Data, **Albert S. Hoagland.** *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 381–85.

11. Ferromagnetism (book), **R. M. Bozorth.** D. Van Nostrand Company, Inc., Princeton, N. J., 1951, pp. 350, 381.

12. Studies on Magnetic Recording (4 parts), **W. K. Westmijze.** *Philips Research Reports*, Eindhoven, Netherlands, vol. 8, 1953.

13. Introduction to Solid State Physics (book), **Charles Kittel.** John Wiley & Sons, Inc., New York, N. Y., first edition, 1953, pp. 167, 168.

14. Principles of Electricity (book), **L. Page, N. I. Adams.** D. Van Nostrand Company, Inc., p. 38.

15. The Reproduction of Magnetically Recorded Signals, **R. L. Wallace, Jr.** *Bell System Technical Journal*, New York, N. Y., Oct. 1951, pp. 1145–73.

16. Magnetic Materials for Digital Com-

puter Components, **Norman Menyuk.** *Journal of Applied Physics,* New York, N. Y., vol. 26, 1955, pp. 8, 692 (detailed study).

17. The Reproduction of Signals Recorded on Magnetic Tape, **E. D. Daniel, P. E. Axon.** *Proceedings,* Institution of Electrical Engineers, vol. 100, pt. III, 1953, p. 157.

18. Magnetic Data Recording Theory: Head Design, **Albert S. Hoagland.** *AIEE Transactions,* vol. 75, pt. I, Nov. 1956, pp. 506–12.

19. Alloy Improves Magnetic Recording, **C. W. Lufcy, W. T. Heath.** *Electronics,* New York, N. Y., June 1955.

20. A Magnetic Tape System for Recording and Reproducing Standard FCC Color Television Signals, **H. F. Olsen, W. D. Houghton,** et al. *RCA Review,* Princeton, N. J., vol. 17, Sept. 1956, p. 330.

21. Magnetic Head Has Megacycle Range, **Otto Kornel.** *Electronics,* Nov. 1956.

22. Calculation of the Magnetic Field in the Ferromagnetic Layer of a Magnetic Drum, **Olle Karlqvist.** *Transactions,* Royal Institute of Technology, Stockholm, Sweden, 1954.

23. Magnetic Field Distribution of a Ring Recording Head, **S. J. Begun.** *Audio Engineering,* Dec. 1948, p. 11.

24. Electromagnetism (book), **J. C. Slater, N. H. Frank.** McGraw-Hill Book Company, Inc., New York, N. Y., 1947, p. 109.

Spectrum Analysis of Random Noise Generators

L. G. POLIMEROU
ASSOCIATE MEMBER AIEE

Synopsis: A method is introduced for evaluating the spectrum shape of random noise generators. This method uses only one selective filter and a variable chopper-carrier to select within a wide band, the frequency which is of interest. The random noise at the frequency selected is then squared and averaged over a sufficient length of time. This process being repeated at various frequencies yields the spectrum shape. The spectrum to be examined by the variable-chopper scheme is preshaped by a wide-band filter of known characteristic. The deviations of the output of the selective filter from the wide-band characteristic indicates the deviation of the original spectrum from a flat-shaped spectrum.

THE OBJECTIVE in this particular study was to determine the power spectrum of some type of random noise. For the sake of discussion, assume the existence of a particular type of noise, namely, a continuous stationary time series, i.e., a noise that has statistical properties which are invariant with time. Such noise will have two fundamental characteristics: an amplitude distribution, which specifies the relative probability that a given noise will occur; and the characteristic with which we are concerned, a power spectrum, which specifies the division of the noise power into frequency components.[1] A previous measurement of the first characteristic mentioned shows that the noise measured has an output characterized by a normal or gaussian distribution.

The probability distribution describes the intensity characteristics of the noise but does not describe the frequency characteristics called the power spectral density. The power spectral density may be calculated numerically as outlined by Tukey.[2] Briefly summarized, this procedure is: first, read the time-history record point by point; second, calculate the autocorrelation function of the data sampled by evaluating the integral $\int_0^T f(t)f(t+\tau)dt$ for various discrete values of time lag τ; and third, determine the power spectral density by taking the Fourier cosine transform of the autocorrelation function. However, there are two types of electronic analog equipment which might be used to determine power spectra: one type parallels the numerical process by first determining the autocorrelation function and then taking its transform to get the power spectral density; the other measures the spectrum directly by means of selective filters. A mechanized device providing a scanning electrical filter is described by Smith.[3]

Mode of Operation Analysis

The principle of operation involved in this analysis is based on the noise detection scheme which is used in the actual noise generator. In the noise generator, a frequency shifting procedure is used to produce noise in the frequency region from direct current to 40 cps (cycles per second). This noise is produced by utilizing a 450-cps electromechanical chopper to shift frequencies in a band pass approximately 100 cps wide, centered at 450 cps. The chopper multiplies the noise voltage alternately by plus one and minus one. This multiplication by a square wave results in frequency components consisting of sums and differences of the noise frequencies, and the chopper frequency and its various harmonics.

Similarly, a band of noise is selected in the frequency region of 350 to 550 cps in the noise evaluation set up. If a variable-frequency chopper is made to operate, in one instance, at 810 cps, it will produce frequency components consisting of sums and differences of all the noise frequencies in that band with 810 cps and the harmonics of 810 cps. Furthermore, if a very-narrow-band selective filter centered at 400 cps is used to select all 400-cps components formed as a result of the sums and differences of the noise frequencies and the chopper frequency and its harmonics, then those noise frequencies which would combine with the chopper frequency and its harmonics to yield components at 400 cps are 410 cps, 1,210 cps, 1,220 cps, 2,020 cps, etc. Among these, the only noise frequency available in the band of noise is the component at 410 cps. Therefore, while operating the chopper at a frequency of 810 cps, a 400-cps component will be formed with an intensity proportional to the corresponding frequency component at 410 cps. Furthermore, by varying the frequency of the chopper, in the frequency region at 750 to 950 cps, frequency components will be produced at 400 cps with an intensity which corresponds to the frequency components in the region of 350 to 550 cps respectively. The effect of varying the chopper frequency at discrete intervals throughout the range mentioned is comparable to having a set of parallel selective filters which are capable of passing a narrow band of frequencies at specified increments in the spectrum region to be examined.

Tests and Results

An examination of Fig. 1 shows a simplified block diagram of the noise generator and the measuring apparatus. The thyratron source output is shaped by a wide-band filter and then amplified. The amplifier output goes to a variable

---

Paper 57-701, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted April 19, 1957; made available for printing March 10, 1958.

L. G. Polimerou, formerly with Convair Division, General Dynamics Corporation, is now with Chance Vought Aircraft, Inc., Dallas, Tex.

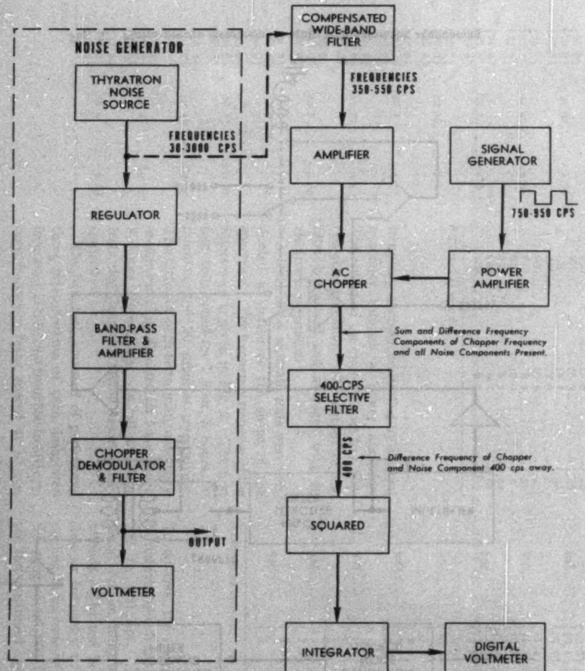

NOISE GENERATOR
THYRATRON NOISE SOURCE
FREQUENCIES 30-3000 CPS
REGULATOR
BAND-PASS FILTER & AMPLIFIER
CHOPPER DEMODULATOR & FILTER
OUTPUT
VOLTMETER
COMPENSATED WIDE-BAND FILTER
FREQUENCIES 350-550 CPS
AMPLIFIER
SIGNAL GENERATOR
750-950 CPS
POWER AMPLIFIER
AC CHOPPER
Sum and Difference Frequency Components of Chopper Frequency and all Noise Components Present.
400-CPS SELECTIVE FILTER
400 CPS
Difference Frequency of Chopper and Noise Component 400 cps away.
SQUARED
INTEGRATOR
DIGITAL VOLTMETER

Fig. 1. Noise generator and measuring setup

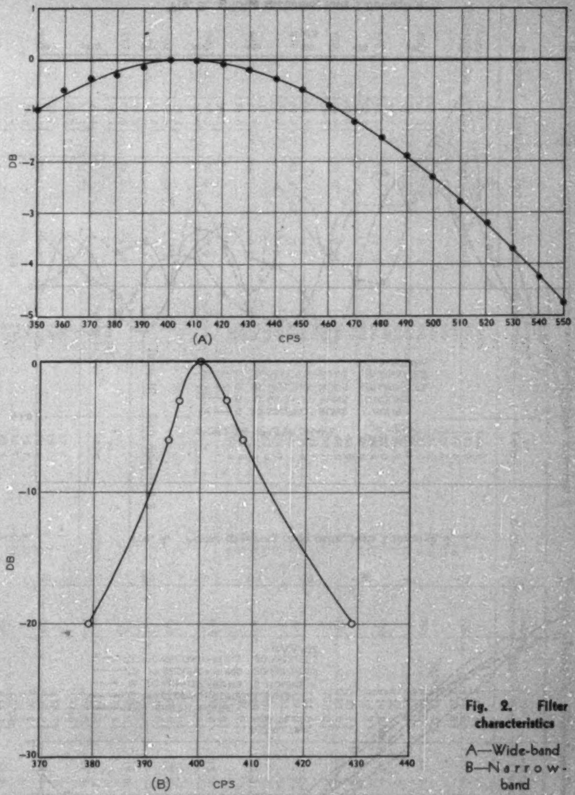


Fig. 2. **Filter characteristics**

A—Wide-band
B—Narrow-band

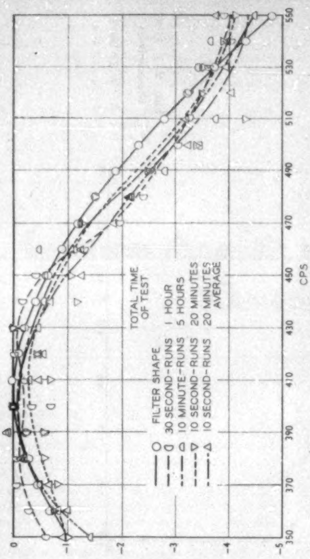


Fig. 4. Noise spectrum composite, tests 1 through 4

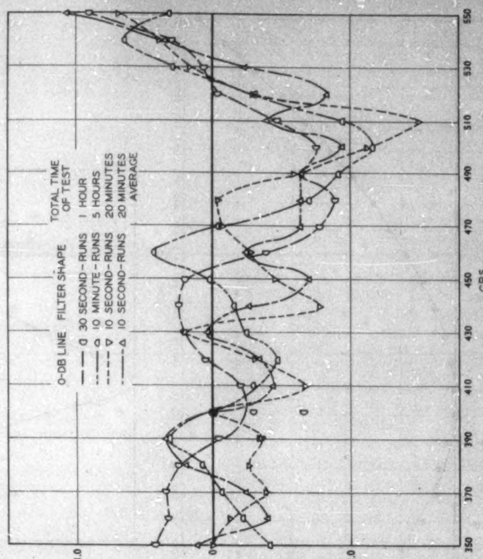


Fig. 5. Noise spectrum, tests 1 through 4

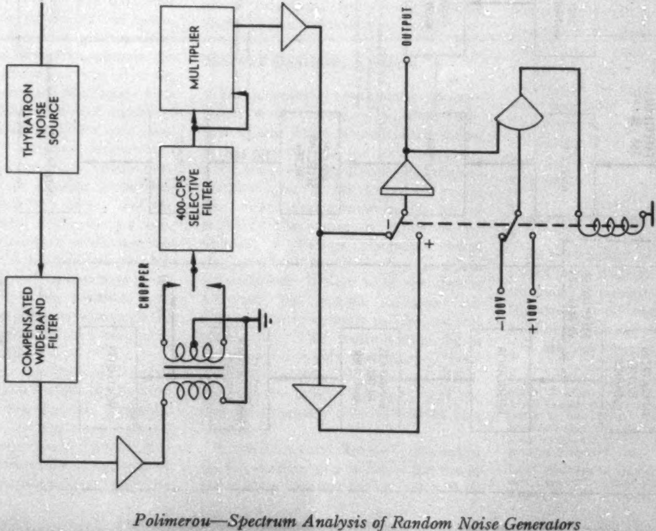

COMPENSATED WIDE-BAND FILTER
THYRATRON NOISE SOURCE
CHOPPER
400-CPS SELECTIVE FILTER
MULTIPLIER
OUTPUT
-
+
-100V
+100V

Fig. 3. Noise-spectra measurement setup with automatic sequencing

Table I. Data from Test 1

Chopper Frequency, Cps	$\overline{v^2}$ Volts	Average of 3 Runs	Decibels
800*	54.4 55.0 56.2	55.2	0
810	54.7 53.7 49.2	52.53	−0.21
820	49.4 56.8 57.4	54.53	−0.05
830	55.2 55.2 55.3	55.23	0
840	54.1 52.8 53.6	53.5	−0.15
850	50.7 52.2 48.0	50.3	−0.4
60	38.8 42.5 41.5	40.93	−1.3
870	35.0 34.6 35.1	34.9	−1.99
0	31.1 31.1 33.0	31.73	−2.4
890	29.7 32.6 29.8	30.7	−2.54
900	26.5 27.5 28.0	27.33	−3.06
910	28.0 25.5 24.3	25.93	−3.28
920	26.0 25.0 27.1	26.03	−3.24
930	25.5 23.2 23.1	23.93	−3.64
940	21.8 21.8 22.4	22.0	−4.0
950	22.7 22.5 22.4	22.53	−3.9
800	46.9 49.7 45.3	47.3	−0.67
800	50.8 52.0 51.3	51.37	−0.32
790	38.0 49.0 48.5	48.5	−0.56
790	49.3 53.7 53.6	52.2	−0.24
780	55.0 54.6 53.8	54.47	−0.05
770	56.2 51.3 55.5	54.33	−0.05
760	52.8 53.0 49.8	51.87	−0.27
750	47.6 50.0 47.3	48.3	−0.58

* Chosen as zero db.

Table II. Data from Test 2

Chopper Frequency, Cps	$\overline{v^2}$ Volts	Decibels
800*	149.5	0
810	139.9	−0.3
820	130.5	−0.58
830	134.5	−0.45
840	131.0	−0.56
850	130.4	−0.58
860	134.1	−0.47
870	112.4	−1.24
880	90.0	−2.2
890	78.8	−2.82
900	67.3	−3.47
910	63.0	−3.76
920	66.8	−3.5
930	68.2	−3.41
940	64.4	−3.66
950	53.4	−4.48
800	149.1	0.0
790	154.7	0.12
780	141.9	−0.23
770	134.3	−0.46
760	123.8	−0.82
750	107.9	−1.42

* Chosen as zero db.

Table III. Data from Test 3

Chopper Frequency, Cps	$\overline{v^2}$ Volts	Decibels
750	17.8	−1.0
760	19.0	−0.72
770	18.6	−0.80
780	19.7	−0.56
790	19.7	−0.56
800*	22.4	0
810	19.2	−0.67
820	20.3	−0.42
830	22.4	0
840	17.8	−1.18
850	17.5	−1.06
860	17.8	−1.18
870	16.7	−1.28
880	15.7	−1.54
890	12.6	−2.5
900	10.2	−3.42
910	8.3	−4.31
920	10.0	−3.5
930	9.9	−3.54
940	9.1	−3.91
950	8.7	−4.12

* Chosen as zero db.

frequency chopper, controlled by a signal generator through a power amplifier. The output of the chopper goes to a 400-cps selective filter. The output of the filter is squared and averaged with an electronic integrator.

A total of six separate tests were conducted to determine the shape of the spectrum. The frequency response of the filters used is given in Fig. 2. The first four tests lasted for 1 hour, 5 hours, 20 minutes, and 20 minutes respectively. The fifth and sixth tests, using automatic sequencing (Fig. 3) lasted for 5 hours and 8 hours respectively. In all of these tests, the integration period was manually controlled.

The first test consisted of performing three measurements at each designated frequency in the band from 350 to 550 cps, each frequency being spaced at 10 cps.

Table IV. Data from Test 4

Chopper Frequency, Cps	$\overline{v^2}$ Volts	Average of 3 runs	Decibels
750	17.3 16.9 15.6	16.6	−0.88
760	16.4 17.0 15.4	16.1	−1.0
770	16.7 19.5 16.3	17.5	−0.64
780	20.1 21.3 18.0	19.8	−0.11
790	21.0 21.0 21.0	21.0	+0.15
800*	20.0 19.8 21.0	20.3	0
810	19.6 16.8 18.4	18.3	−0.45
820	18.0 18.0 19.0	18.3	−0.45
830	20.2 21.3 17.4	19.6	−0.16
840	18.2 18.6 15.5	17.4	−0.67
850	19.1 13.3 16.0	15.1	−1.29
860	17.4 15.3 13.2	15.3	−1.23
870	13.9 12.5 13.5	13.3	−1.84
880	12.0 12.1 13.1	12.4	−2.14
890	11.4 11.7 10.9	11.3	−2.54
900	10.2 9.4 9.3	9.6	−3.25
910	9.9 10.2 9.1	9.7	−3.2
920	8.8 7.2 7.9	8.0	−4.04
930	8.5 9.4 7.2	8.2	−3.94
940	8.6 8.6 7.5	8.2	−3.94
950	8.0 10.9 7.0	8.6	−3.73

* Chosen as zero db.

To do this, the chopper frequency was varied correspondingly over the range from 750 to 950 cps. Each measurement lasted for an integration period of 30 seconds. The test began by setting the chopper frequency to 800 cps, and proceeding in increments of 10 cps to cover the range from 800 to 950 cps. Then, the test continued again at 800 cps, and

Table V. Data from Test 5: Seconds Per Cycle Per 5

8.4	10.3	9.7	9.5	9.4	10.0	9.0	8.5
8.6	10.6	9.4	8.5	9.3	9.3	9.5	9.3
9.1	10.2	9.2	9.3	9.6	9.9	9.0	8.4
9.9	10.0	9.8	9.1	9.1	9.4	9.3	8.9
8.8	10.0	10.0	9.2	9.3	9.6	9.7	8.9
9.3	10.3	10.0	9.2	9.5	9.5	9.8	9.0
9.0	10.0	9.7	9.3	9.3	9.9	9.7	8.3
9.5	9.5	10.0	9.3	9.5	9.8	10.0	8.6
10.5	9.7	10.3	9.2	10.0	9.7	9.6	8.3
10.7	10.1	9.6	9.9	9.8	9.3	9.5	8.3
10.0	10.3	10.2	9.1	9.7	9.7	9.0	8.0
10.0	9.6	9.5	9.4	9.5	9.8	9.6	8.0
9.8	9.6	9.1	9.4	9.8	9.7	9.4	8.9
9.1	9.2	9.6	9.5	9.3	9.5	9.5	7.9
9.2	8.1	9.2	9.2	9.1	9.5	10.3	8.3
9.2	8.6	9.1	9.6	9.0	9.3	10.5	8.0
9.3	8.3	9.3	9.3	9.5	8.4	9.8	8.5
8.8	9.0	9.2	9.5	9.4	9.3	9.5	8.2
9.0	9.0	9.7	9.4	9.5	9.5	9.8	8.0
9.4	9.7	9.3	9.4	9.4	9.6	9.4	8.1
9.4	9.1	9.8	9.6	9.5	9.7	8.8	7.9
8.5	9.5	9.7	9.7	9.9	9.0	9.3	8.2
9.0	8.7	9.6	9.7	9.7	9.3	9.2	8.7
9.4	8.8	9.6	9.5	10.0	9.0	9.2	9.0
8.8	8.4	9.9	10.1	10.3	9.4	8.8	9.0
9.2	8.9	10.0	9.9	10.2	8.5	9.0	9.3
9.4	9.8	10.2	9.7	10.3	8.7	9.0	9.2
9.0	9.7	10.3	10.1	10.3	9.4	8.8	9.0
9.7	9.3	9.7	9.6	10.1	9.4	8.6	8.9
9.7	9.7	9.6	9.6	10.2	9.4	8.8	8.8
9.8	10.1	9.4	10.2	10.3	9.0	9.2	9.0
9.9	10.0	10.7	9.3	10.0	9.3	9.5	9.0
9.2	10.4	10.2	9.9	9.6	9.6	9.1	9.3
9.0	9.7	10.0	9.8	9.8	9.0	8.3	9.3
9.5	10.3	10.0	10.0	10.2	8.2	9.4	9.0
9.2	10.2	9.8	10.0	10.3	9.0	9.0	8.8
9.5	10.8	9.1	9.8	10.3	9.2	8.9	8.9
9.4	10.8	8.8	10.0	10.1	9.4	8.7	9.1
9.6	9.9	9.2	10.0	10.1	9.5	8.4	8.7
9.3	10.3	9.1	9.3	9.8	9.1	8.2	8.9
9.6	9.6	8.3	10.0	9.3	9.0	8.3	8.7
8.8	9.7	7.8	9.7	10.0	9.1	8.3	8.8
9.7	9.4	8.5	9.6	10.0	9.3	8.3	8.5
9.9	9.9	8.3	9.6	10.0	9.5	8.5	8.8
9.7	9.3	8.9	9.7	10.0	9.4	9.2	9.0
						8.4	9.2
Total							3395.7
Average							9.38

Note: 3395.7×5 total time elapsed for 362 cycles from −100 to +100 to −100 volts or 362×400 volts = 140,800 volts.

Table VI. Frequency Chart from Test 5

x_i, Seconds Per Cycles Per 5	f_i	f_ix_i	$f_ix_i^2$	x_i^2
7.8	1	7.8	60.84	60.84
7.9	2	15.8	124.82	62.41
8.0	4	32.0	256.00	64.00
8.1	2	16.2	131.22	65.61
8.2	4	32.8	268.96	67.24
8.3	11	91.3	757 79	68.89
8.4	6	50.4	423.36	70.56
8.5	8	68.0	578.00	72.25
8.6	4	34.4	295.84	73.96
8.7	6	52.2	454.14	75.69
8.8	14	123.2	1084.16	77.44
8.9	9	80.1	712.89	79.21
9.0	28	252.0	2268.00	81.00
9.1	15	136.5	1242.15	82.81
9.2	21	193.2	1777.44	84.64
9.3	31	288.3	2681.19	86.49
9.4	25	235.0	2209.00	88.36
9.5	27	256.5	2436.75	90.25
9.6	23	220.8	2119.68	92.16
9.7	28	271.6	2634.52	94.09
9.8	17	166.6	1632.68	96.04
9.9	12	118.8	1176.12	98.01
10.0	26	260.0	2600.00	100.00
10.1	7	70.7	714.07	102.01
10.2	9	91.8	936.36	104.04
10.3	14	144.2	1485.26	106.09
10.4	1	10.4	108.16	108.16
10.5	2	21.0	220.50	110.25
10.6	1	10.6	112.36	112.36
10.7	2	21.4	228.98	114.49
10.8	2	21.6	233.28	116.64
	362	3395.2	31964.52	

$$s^2=\frac{\Sigma fixi^2-\dfrac{(\Sigma fixi)^2}{n}}{n-1}$$

$s^2=0.335$

$s=0.58$

$M+2s=9.38\pm1.16$ seconds per cycle per 5

Table VII. Cumulative Distribution Data from Test 5

x_i, Seconds Per Cycles Per 5	f_i	Per Cent
8.2	4	1.18
8.3	15	4.44
8.4	21	6.21
8.5	29	8.58
8.6	33	9.76
8.7	39	11.54
8.8	53	15.68
8.9	62	18.34
9.0	90	26.63
9.1	105	31.06
9.2	126	37.28
9.3	157	46.45
9.4	182	53.85
9.5	209	61.83
9.6	232	68.64
9.7	250	73.96
9.8	267	78.99
9.9	279	82.54
10.0	305	90.24
10.1	312	92.31
10.2	321	94.97
10.3	335	99.11
10.4	336	99.41
10.5	338	

similarly proceeded in increments of 10 cps to cover the range from 800 to 750 cps. The average value of $\overline{v^2}$ was computed from each set of three values at each designated frequency. The total time was one hour. A plot of the spectrum shape is then constructed showing the faired curves. The results of the first four tests are plotted in Figs. 4 and 5. Tables I through VIII contain data of all the tests, and Fig. 6 gives the cumulative distribution of test 5.

The second test was repeated at the designated frequencies in the same manner. However, each integration period lasted for 10 minutes, and only one measurement was conducted at each designated frequency. The total time was 5 hours.

Again, for the third test, only one measurement was conducted at each designated frequency. The integration period lasted for 30 seconds for each measurement. The total time was 20 minutes.

On the fourth test, three measurements were made at each designated frequency. The integration period lasted for 10 seconds each. The total time was 20 minutes.

The fifth test consisted of choosing only one frequency and using an integration period of 5 hours. To be able to do this, an automatic sequencing technique was utilized. This sequencing involved reversing the sign of the information at the input to the integrator whenever the output of the integrator reached ±100 volts. The output of the integrator is made to cycle between the limits of −100 volts to +100 volts. Variations in the period required to complete such a cycle were recorded to determine any nonstationary effects at the particular frequency. If it is desired, a long-time average of the energy in a narrow band of that frequency is also made available. In this case, a voltage of 140,000 volts was reached corresponding to the total number of sequencing cycles obtained at a frequency band centered at 450 cps in the noise output. Results of this test are found in Table V.

The sixth test consisted of varying the chopper frequency over the region from 800 to 900 cps in ten cps steps. The arrangement of the frequencies measured was constructed to represent the spectrum at three different intervals of time. During the first part of this test, $\overline{v^2}$ was evaluated only at the following frequencies: 400, 420, 450, 480, and 500 cps. The second part evaluated the following

Table VIII. Data from Test 6

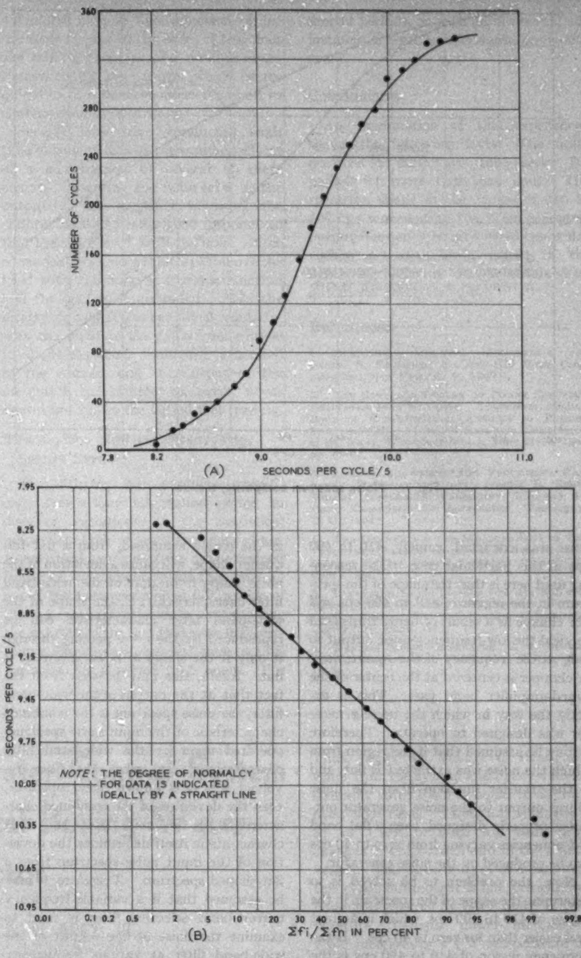
Chopper Frequency, Cps	Sequence Cycles, $\overline{v}^2$ per 400	Decibels
800*	77.75	0
810†	75.90	0
820*	69.70	−0.48
830†	73.40	−0.37
840‡	71.15	−0.28
850*	42.35	−2.64
860‡	53.00	−1.56
870†	47.65	−2.02
880*	39.20	−3.00
890†	39.75	−2.81
900*	35.10	−3.46

* First set of frequencies tested.

† Second set of frequencies tested.

‡ Third set of frequencies tested.

Note: 800 cps chosen as zero db.



NOTE: THE DEGREE OF NORMALCY FOR DATA IS INDICATED IDEALLY BY A STRAIGHT LINE

Fig. 6. Cumulative distribution from test 5. A—Frequency graph. B—Test for normalcy; see Tables V, VI, and VII

frequencies: 410, 430, 470, and 490 cps. Also, the last part included the remaining frequencies: 440 cps and 460 cps. The results of the test are plotted in Fig. 7.

Discussion

In measuring the output of the noise generator which is being evaluated, certain fundamental limitations exist. Evaluation of the spectrum in its original region provides an easier analysis. An analysis of errors present is important. The type of noise being evaluated is also of interest.

FUNDAMENTAL LIMITATIONS

Certain fundamental limitations exist which necessitate determining the length of time necessary to establish accurate estimates of the statistical properties of the noise. Determining the spectrum shape at arbitrarily low frequencies requires an observation time which may become prohibitively long. Similarly, as the resolution of the spectrum becomes arbitrarily fine, the observation time again may become prohibitively long. This is a fundamental limitation and not a limitation imposed by technique. Bennett and Fulton[4] show the derivation of the equations for the length of time necessary to establish satisfactory estimates of the statistical properties of low-frequency noise. For instance, an equation which gives the time required to insure with 95% certainty that the mean value of the noise lies within $P\%$ of the true mean is

$$T=(2\times 10^4 N_0)/P^2\bar{e}^2 \text{ seconds}$$

where T is the time required in seconds, N_0 is the spectral density in volt² per cps, and $\bar{e}$ is the true mean value in volts. For example, to determine within an accuracy of 10% the mean value of the output of the noise generator which has a true mean value $\bar{e}$ of 0.05 volt and a spectral density, N_0, of 0.6 volt² per cps would require

$$T=\frac{2\times 10^4\times 0.6}{(10)^2\times(0.05)^2}=4.8\times 10^4 \text{ seconds}$$

or over 13 hours of integration time. Similarly, the time required to evaluate the spectral density with the use of an ideal narrow-band filter, i.e., zero attenuation for $f_0<f<f_0+\Delta f$ and infinite attenuation for all other frequencies, is shown to be approximately $T=(4\times 10^4)/P^2\Delta f$ seconds, provided $f_0>>\Delta f$.

MEASURING ORIGINAL SPECTRUM

To measure the spectrum shape of the noise output, and avoid lengthy observation periods, it is convenient in many instances to measure the region of the spectrum from which the low-frequency

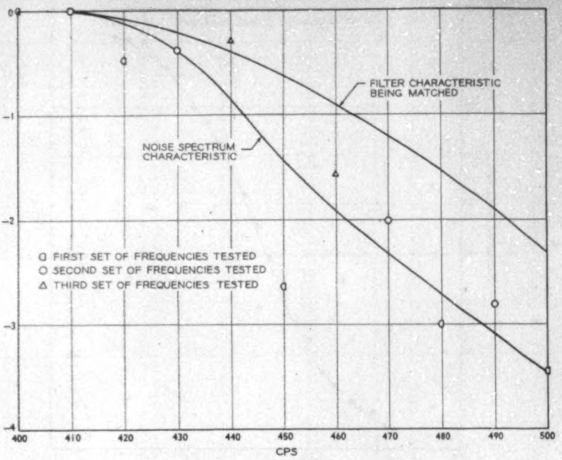


Fig. 7. Noise spectrum composite, test no. 6

noise was extracted, namely, 410 to 490 cps in this particular case. The reasoning used here is that the shape of the spectrum in the region of 410 to 490 cps will not change as a result of being translated to yield the low-frequency noise output so long as the frequency of the square wave or chopper is centered at the center of the tuned-amplifier band pass. This is exactly the way in which the noise generator was designed to operate. Therefore it may be assumed that if the region from which the noise was extracted is flat, and if the transfer function from the noise source output to the noise generator output is properly designed, then a flat band of frequencies varying from zero to 40 cps can be produced by the noise generator.

Now, the problem to be solved is to determine the shape of the spectrum in the region of 410 to 490 cps, which is a great deal easier than for zero to 40 cps. If the frequency region of 410 to 490 cps is the region of interest, a region somewhat larger than this must be examined so that edge effects (effects which result from the fact that the narrow-band selective filter meausres more than just a narrow frequency band centered at the frequency being measured) are lessened. In this set-up, a region over twice as large as the region of interest was selected as that to be examined for most of the experiment, namely, the frequency region from 350 to 550 cps. Having determined the wide-band filter characteristic, Fig. 2, any deviation of the noise source output in the region examined, from a flat frequency shape, will cause a deviation of the noise shape from that of the wide-band filter characteristic, i.e., the shape of the wide-band filter characteristic can be reproduced by the noise passing through it only if the shape of noise spectrum is flat. Again, this may be seen from the fact that at the output of the wide-band filter, the noise spectrum is the combined characteristic of the input noise spectrum operated upon by the wide-band filter characteristic, and since the filter involved acts as a known linear operator, then the deviation of the combined characteristic of the noise from the filter characteristic itself determines the deviation of the input noise spectrum from a flat-shaped spectrum. Therefore, it may be assumed that if a variable-frequency narrow-band selective filter is used to examine the noise at the output of the wide-band filter at various frequencies, then the output of the narrow-band selective filter will reproduce the wide-band frequency characteristic only if the noise input to the wide-band filter is flat-shaped. Furthermore, the deviation from reproducing the wide-band filter characteristic is the same deviation of the wide-band filter input noise shape from a flat spectrum.

With this basic material in mind, it remains to translate the frequency component which is to be measured with a single fixed-tuned narrow-band selective filter. Since, as indicated earlier, the process of chopping an electric noise signal by using an electromechanical chopper produces sums and differences of all frequency components present in the noise with the chopper frequency and its harmonics, then by limiting the frequencies to be examined, linear detection can be obtained. Also, a frequency component can be produced at 400 cps, proportional in intensity to any frequency component to be examined, by varying the chopper frequency, in such a way that its frequency minus 400 cps results in the frequency component to be measured. Once more, this is to say that a frequency component is produced at 400 cps which is proportional to the frequency component f_i being examined, f_c (chopper-frequency) minus f_i produces the 400-cps component. Also, since no other frequency in the noise range being examined, except f_i, produces a 400-cps frequency component with any one particular f_c, this detection scheme is essentially linear.

TYPES OF ERRORS

There are many possible sources of error in an experiment of this type, and each must be examined and minimized. Among these sources of error may be included arcing of the chopper contacts and contact bounce, pickup noise in cabling due to stray fields, drift in d-c amplifiers and integrators, level of signals sent to multipliers being limited, and noise originating in the multipliers and nonlinearities introduced by squaring the noise in the multipliers, and errors due to manual timing procedures.

To counteract the first source of error the noise voltage was amplified before reaching the chopper and a small-valued capacitor was used to minimize arcing. Also, the chopper driving-voltage level and shape must be controlled to give minimum bounce. Another factor is obtaining the proper matching impedance for the transformer used to convert from a single-ended amplifier to the balanced operation of chopping.

To prevent adulteration of the noise from stray fields, the voltage level was amplified to a high level before sending the noise signal over wires of any length. Also, the properly shielded wires were used when sending the noise voltage over the 100-foot cable to the computer console and from the console to the diode-electronic multiplier.

Drift was combatted in the amplifiers by using a minimum number of selected amplifiers. Drift in the integrator was reduced by using one of several specially converted integrators giving extremely low drift over periods in terms of hours. Also, by the use of automatic sequencing, alternating the direction of integration tended to average out such drift.

Also, as the frequency of the signals supplied to d-c amplifiers increased, limiting occurred due to output capacity of the d-c amplifier, which meant that voltage levels from the amplifiers could not exceed 30 volts without overloading. This in turn meant that peak voltages of more than 9 volts could not be obtained from the multipliers, since the multipliers produce an output voltage XY/100. As a result, since the signal output of multipliers is limited to low voltages, the accuracy involved is highly sensitive to drift and errors due to nonlinearities of the multiplier. However, again if automatic sequencing is used, even these errors are diminished. Also, where trouble in squaring high-frequency noise is involved, another means is available under certain conditions. If the noise is normally distributed and the mean value of the noise can be eliminated, then the mean-square value of the noise may be inferred from the average of the positive (or negative) values of the noise voltage according to the equation $\overline{v^2}=2A^{+2}$, where A^+ is the average value of the positive values of the noise.[4] A^+ can be conveniently determined by rectifying the noise and then integrating. This method has excellent frequency response and can yield considerable accuracy.

Finally, manual determination of increments of time in the order of less than one minute can also add notable error. Precise timing mechanisms would be required to turn the computer on and off to eliminate this type of error.

Several tests were conducted from time to time to look at the summed effects of a combination of several of these errors. Observing the noise level at the output of the amplifier following the chopper showed a fairly good response, so that the noise level, with the noise source connected, was more than 20 db above the level with the noise source disconnected and the lead to it grounded. Also, the integrator output was examined regularly, with the input to the cable from console to diode-electronic multiplier grounded at the console, and the output at the integrator represented an error never more than 2% of the actual test reading.

TESTING FOR NONSTATIONARY TIME SERIES EFFECTS

The fifth test was conducted specifically over a long integration period to discover variations of the statistical properties of the noise with respect to time. A statistical analysis of the data obtained from this experiment shows that only 14 out of 362 measured values fall outside of the $\pm 2\sigma$ region for this data, thus showing the validity of the fact that with 95% certainty, the data shows less than ±13% variation for a period of 5 hours. For shorter periods, or with an occasional adjustment of level, smaller variations would result.

Conclusions

An examination of this experiment shows the following facts: The noise presents a stationary time series for periods of more than one hour. The spectrum shape in the region of 410 to 490 cps representing the noise generator output frequency of zero to 40 cps is flat within ±1 db. Improvement in the measuring technique may be obtained by automatic timing circuits.

References

1. FREQUENCY ANALYSIS, MODULATION AND NOISE, S. Goldman. McGraw-Hill Book Company Inc., New York, N. Y., 1948.

2. THE SAMPLING THEORY OF POWER SPECTRUM ESTIMATES, John W. Tukey. *Symposium*, Applications of Autocorrelation Analysis to Physical Problems, Office of Naval Research, Department of the Navy, Woodshole, Mass., June 13–14, 1949, pp. 46–67.

3. ANALOG EQUIPMENT FOR PROCESSING RANDOMLY FLUCTUATING DATA, Francis B. Smith. Langley Aeronautical Laboratory, National Advisory Committee for Aeronautics, Washington, D. C., 1954.

4. THE GENERATION AND MEASUREMENT OF LOW FREQUENCY RANDOM NOISE, R. R. Bennett, A. S. Fulton. Hughes Aircraft Company, Culver City, Calif., 1954.

5. STABILIZED NOISE SOURCE FOR AIR-WEAPONS DESIGN, D. E. Beecher, R. R. Bennett, H. Low. Hughes Research and Development Laboratories, Culver City, Calif., 1954.

6. RANDOM NOISE GENERATOR, C. S. Deering. *TPN-23*, Convair-Fort Worth, Fort Worth, Tex., 1955.

An Electronic Feedwater Control System for the Experimental Boiling Water Reactor

R. H. MORSE
NONMEMBER AIEE

R. N. BREY, JR.
NONMEMBER AIEE

THE feedwater (FW) system for the Experimental Boiling Water Reactor was designed for utmost flexibility to accommodate many different experiments and procedures scheduled for this facility now in operation at the Argonne National Laboratory. Design criteria called for control of level from either the hotwell or the reactor vessel and ability to switch from simple single-element control (level control adjusting the main FW valve) to 2-element control, using either steam flow (SF) or FWF (FW flow) as the second element, to 3-element control using level, FWF and SF. Further, all of these functions had to be adjustable as to their worth, one to the other, and the final resultant signal to the FW control valve had to be adjustable as to proportional band, reset (integral), and rate action (derivative), to allow 3-function control.

As may be seen from the over-all plant flow diagram in Fig. 1, the steam cycle is a simple single cycle enclosed with two main storage capacities for water, i.e., the condenser hotwell and the reactor vessel. Thus, neglecting leakage of steam and water from the system and under a steady steaming rate, FWF is equal to SF on a mass flow basis. Thus, any changes in SF or FWF or both will cause level changes in the hotwell and in the reactor vessel. The power output of a reactor of this type, assuming it is hot, critical, and boiling, is affected by FWF and FW temperature. Reactor power increases if FWF increases or if FW temperature decreases, thus a method of limiting the speed with which the FWF water can be changed is desirable since any large fast flow changes can have serious transient effects on reactor power. An electric valve drive mechanism has this desirable

---

Paper 58-700, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Chicago, Ill., March 17–21, 1958. Manuscript submitted December 23, 1957; made available for printing March 10, 1958.

R. H. MORSE and R. N. BREY, JR., are with the Leeds & Northrup Company, Philadelphia, Pa.

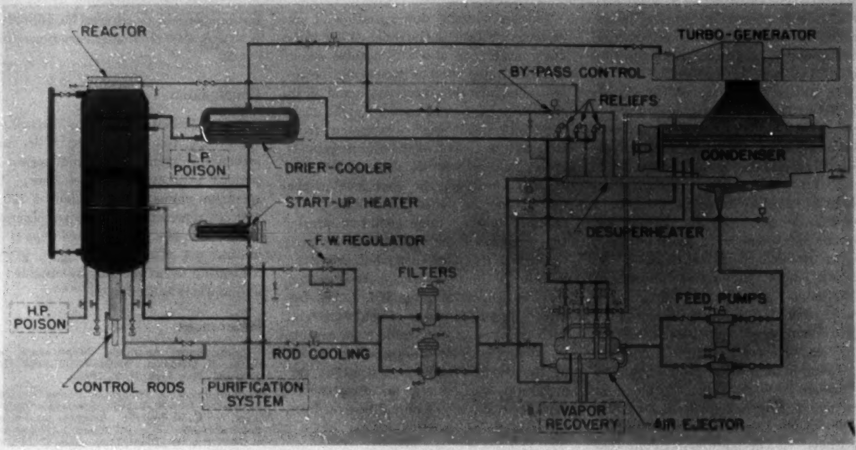

REACTOR
TURBO-GENERATOR
BY-PASS CONTROL
RELIEFS
L.P. POISON
DRIER-COOLER
CONDENSER
START-UP HEATER
DESUPERHEATER
F.W. REGULATOR
FILTERS
FEED PUMPS
H.P. POISON
ROD COOLING
CONTROL RODS
PURIFICATION SYSTEM
VAPOR RECOVERY
AIR EJECTOR

Fig. 1. Plant flow diagram (courtesy of Argonne National Laboratory) of the single-cycle experimental boiling water reactor at Argonne National Laboratory

feature plus adjustability of maximum *dw/dt* that can be introduced.

Rapid and large changes in level in the reactor vessel can also affect reactivity and thus power level by temporarily affecting bubble formation and the boiling heat transfer coefficient due to the head change, thus it is desirable to hold level as close as possible to set point even during large changes in steaming rate without large fluctuations in FWF. In order to minimize upsets to reactor level, this system was designed so that the SF response action can be delayed to compensate for the effects of "shrink" and "swell" which take place during load changes. This effect is inherent, either in conventional boilers or in boiling water reactors. As steaming rate is increased the additional steam bubble formation replaces saturated water in the vessel. This causes decreased over-all density and a rise in level in the vessel. The reverse is present when steaming rate is decreased. To appreciate this fully, consider the water within the reactor vessel to be at saturation pressure and at normal steaming level but with no steam being formed. Now consider the formation of steam bubbles around the core. It becomes apparent that for each steaming rate a different weight of water is present in the reactor vessel, assuming the level is held constant. In order to maintain this constant level, on an increase of steaming rate, an amount of water must actually be removed from the reactor vessel as steam. On a decrease in steaming rate, water must be added.

The SF delay network was designed with proper flexibility so that on changes in steaming rate the water flow could be maintained constant for a period of time and then moved to the proper value, or the water flow could be temporarily moved in the opposite direction to SF, depending on which adjustment tends to produce the maximum of stability in the process. The SF delay, therefore, permits the control action to decrease water flow momentarily on an increase in load and increase water flow momentarily or a decrease in load.

The system as designed appeared to require the considerable complexity which is present. This was caused by the number of unknowns in a project of this nature. Future systems may be simplified by substitution of valve position for FWF, and megawatts, or some other readily available measure of output, as an antici-

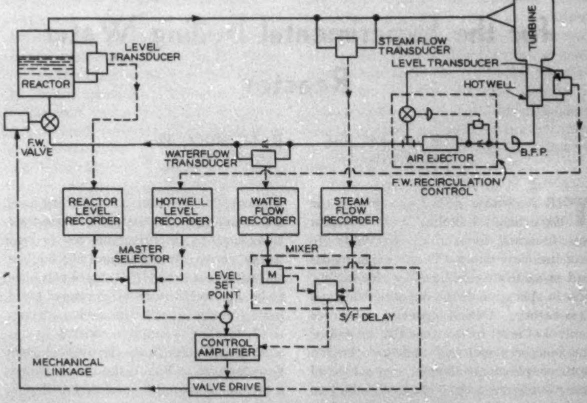

TURBINE
LEVEL TRANSDUCER
REACTOR
STEAM FLOW TRANSDUCER
LEVEL TRANSDUCER
HOTWELL
F.W. VALVE
WATERFLOW TRANSDUCER
AIR EJECTOR
B.F.P.
F.W. RECIRCULATION CONTROL
REACTOR LEVEL RECORDER
HOTWELL LEVEL RECORDER
WATER FLOW RECORDER
STEAM FLOW RECORDER
MASTER SELECTOR
MIXER
M
LEVEL SET POINT
S/F DELAY
CONTROL AMPLIFIER
MECHANICAL LINKAGE
VALVE DRIVE

Fig. 2. Block diagram of FW control system

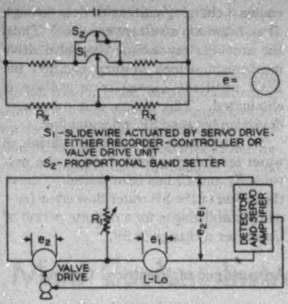

S2
S1
e=
Rx
Rx
S1-SLIDEWIRE ACTUATED BY SERVO DRIVE. EITHER RECORDER-CONTROLLER OR VALVE DRIVE UNIT
S2-PROPORTIONAL BAND SETTER
R1
e2
e1
e2-e1
DETECTOR AND SERVO AMPLIFIER
VALVE DRIVE
L-Lo

Fig. 3. Single-element proportional position control system

pating signal rather than SF-FWF ratio. These modifications will leave the basic system unaltered but employ simplified measuring techniques.

Components of the System

The major components of the control system, as seen in Fig. 2, are electronic SF, FWF, reactor level, and hotwell level transducers; SF, FWF, reactor level, and hotwell level Speedomax recorders; mixer; servoamplifier; and electric FW valve drive mechanism and valve assembly.

Transducers

The transducers for flow and level are differential-pressure-actuated electronic units. The meter body unit converts the linear mechanical motion of stainless-steel differential bellows into a rotary-motion torque-tube output. This torque-tube motion displaces a balanced beam in the transmitter unit, changing current in an electronic oscillator circuit. Current changes of 0.5 to 5 milliamperes are employed in both the output and the feedback (force balance).

Recorders

The Speedomax recorders accept the current output from the transducers by looking at a 50-millivolt d-c drop across a precision 10-ohm resistor. Since the transducers have an output proportional to differential pressure and since differential pressure varies as the square of the flow rate in pounds per hour, it is necessary to employ circuitry in the flow recorders which extracts the square root of the input and automatically adjusts recorder amplifier gain in order to have a linear scale and to retransmit linear intelligence to the control network. (This network computes the action to be taken from the linear signals received and is designated as the mixer.)

The control point as indicated by a red pointer on the level recorder can be varied over the 40-inch range of the recorder scale by means of a control point setter knob. The range is calibrated −20 to 0 to +20 inches of water. The control system then functions to keep the level of the reactor vessel or hotwell, as selected, at the control point setting. Since the total span of possible level measurement and control in the reactor vessel is 96 inches, a coarse zero setter is provided on the reactor vessel level recorder which permits adjustment of the zero point of the recorder between 20 and 76 inches of level.

Mixer

The mixer includes the necessary circuitry to provide the required control action from either the error signal developed between water level and its set point ($L-L_0$) and a signal proportional to FWF-SF ratio. Included in this mixer is a delay network which times the FWF-SF ratio response to compensate for swell and shrink on load changes. This mixer also provides control functions of proportional, reset, and rate actions. To provide complete flexibility, each function is adjustable over a wide range. In the event that any of these functions are not desired, they may be cut in and out by means of selector switches. The unit contains the following control adjustments:

SF response cut-out switch.
Water flow response cut-out switch.
Master level control selector switch.
SF proportional band adjust.
SF time-delay adjust.
Water flow response-time adjust.
Valve drive proportional band adjust.

As mentioned in the foregoing, the intelligence to the mixer unit originates from bridge circuitry which contains retransmitting slide wires from the FWF, SF, and level recorders and the FW valve drive mechanism.

When the control is in full 3-element operation, the retransmitting slide wire on the valve drive unit output shaft provides an instantaneous velocity feedback which is progressively "washed out" to be replaced by the exactly metered water flow feedback. This system, therefore, compensates for process and metering lags and any nonlinearity in the valve characteristic.

Control Amplifier

The control amplifier can be considered as consisting of three basic parts; the converter, the voltage amplifier, and the relay section. It is essentially a polarity-sensitive detecting element. Thus, a d-c signal of a given polarity resulting from mixer unit unbalance will, when processed, energize the relay coil and actuate the valve drive motor with its accompanying feedback slide wire to drive the bridge in the direction of rebalance. Two relays are provided in the amplifier, one operating from a negative signal and the second from a positive signal.

Motor Valve Drive Unit and FW Valve

The final control element in the system is an electrically operated FW valve with stainless-steel body and internals. Stainless steel is used to prevent corrosion. The valve is sized for correct throttling action and internals are selected to characterize the valve action as nearly as possible to give a straight-line relationship of flow rate versus valve stem position. A special leak-off connection is supplied to collect any leakage from the packing when the plant is operated with heavy water.

A mechanical linkage converts the rotary motion of the electric drive unit to a linear motion to drive the valve stem. Further adjustments can be made in this linkage arrangement to characterize the valve under plant operating conditions.

The drive unit itself has two worm-gear reduction stages driven by a 3-phase reversible motor. Position transmitting slide wires and travel limit switches are included in the assembly. The double-reduction worm-gear arrangement provides a self-locking feature which prevents the unit from being "back driven" when de-energized. The position-indicating slide wires provide indication of true valve position of the control panel and feedback to the control network.

Operation of the System

Single-Element Control

Fig. 3 outlines the simple proportional position control circuitry where the FW valve position is maintained proportional to reactor level. Note that in the bridge circuitry above deviation voltage e is designated by a circle which is used in all following figures for purposes of simplicity. S_1 represents a slide wire mounted in the recorder or drive unit. Any motion of the slide-wire contact from its mid-position produces an error voltage.

The control system compares the error signal from the level deviation circuit e_1 and drive unit slide wire e_2, pro-

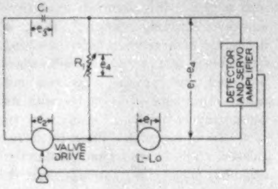


Fig. 4. Single-element proportional plus reset action control

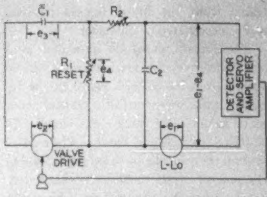


Fig. 5. Single-element proportional plus reset plus rate action control

ducing a corresponding error signal (e_1-e_2) to the detector and amplifier which will result in repositioning the drive unit until (e_1-e_2) is zero.

The relative travel of the drive unit with respect to the level error is the proportional band of the system. This is readily adjustable in the bridge circuitry by the relative settings of S_2.

Fig. 4 indicates the addition of reset action to this circuitry by a simple modification. This comprises the addition of capacitor C_1. This capacitor makes it impossible for the control bridge to be balanced when there is any level deviation e_1 except by a slow and continuous movement of the drive unit e_2 which produces a charging current on capacitor C_1 and a voltage drop e_4 on R_1.

The time constant of network R_1C_1 is adjusted to match the response time of the process and produces a rapid and steady return to control point on any upset in level.

The reset speed is proportional to deviation since as e_1 increases a larger charging current on C_1 results because e_4 must be larger. However, for a given level deviation the speed of valve action remains constant since the rate of increase of e_3 is constant because $E=Q/C$.

Rate action is available although not generally required on this porcess. Rate action operates to add a further increment of valve action as a function of the rate of change of level.

This action is provided by the addition of R_2 and C_2 as shown in Fig. 5. Rate action is produced if the rebalancing voltage e_4 is attenuated.

TWO-ELEMENT CONTROL

Fig. 6 illustrates the substitution of FWF for valve position as the output of the circuit, retaining valve position as a velocity feedback to compensate for process and metering lags and nonlinearities in the process.

An error signal e_1, representing level deviation, is detected and amplified, moving the valve to a new position. The change in voltage e_2 produces a charging current on C_3 which passes through C_1 and develops a voltage e_4 on R_1, rebalancing the amplifier. As the change in FWF is detected (e_5), the charging current on C_3 decays and is replaced by e_5. The time constant of R_3C_3 is selected to match the process and metering lags encountered. Therefore, the valve position feedback "washes out" and is employed only as a velocity feed back.

The reset action in the circuitry operates as described in the foregoing. A previously mentioned, the foregoing functions are completely adjustable and may be cut out if so desired.

THREE-ELEMENT CONTROL

Fig. 7 illustrates the addition of SF-water flow ratio anticipation to the circuit, with its accompanying delay network.

A voltage e_6 is developed across the bridge by the metered SF. Any deviation between SF and water flow (e_6-e_5) causes a charging current to flow through C_1 and develop a voltage e_4 on R_1. Thus, the amplifier repositions the valve drive mechanism, which in turn changes the FWF until the (e_6-e_5) error voltage is eliminated. This action eliminates any deviation between SF and water flow.

To compensate for swell and shrink of water level during load changes, the network R_4 and C_4 has been added to delay the action of the SF-water flow error (e_6-e_5) by stabilizing e_6 for a definite period of time after a change in SF.

Advantages of Electrical Control

1. *Permanent Calibration.* In the electric system, where resistors and capacitor values can be readily controlled, the control-response dial calibrations can be accurately determined and maintained constant.

2. *Linear Response.* In the electric system, a linear rate of reset is available over the entire valve range.

The same is true also for rate-time calibration, as the electric system is unaffected by compressibility effects.

3. *Readily Serviced.* Electric systems lend themselves to unitized designs in which various units or components can be plugged together or separated quickly for servicing.

4. *Transmission Lag.* Electric transmission is not subject to lag. This is of particular importance on fast-responding processes such as flow control.

5. *Flexibility.* The electric system is extremely flexible from the standpoint of combining other features into a standard control system.

6. *Building Integrity.* In the Experimental Boiling Water Reactor, the all-electric control system adapted itself readily to maintenance of building integrity. The containment philosophy required that a minimum of piping or tubing penetrations be made in the primary plant containment vessel. The rubber insulation was vulcanized to the wires to

Vol.
Pt. 1
195

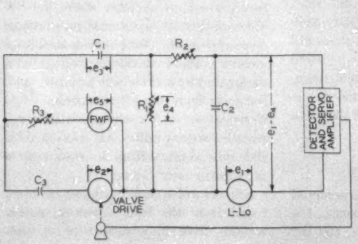
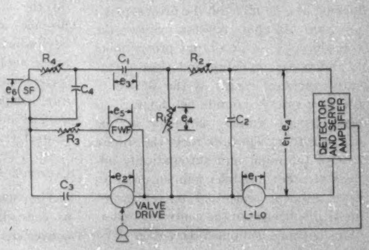


Fig. 6 (left). Two - element proportional plus reset plus rate action control

Fig. 7 (right). Three - element proportional plus reset plus rate action control



prevent gas streaming along the conductors. Wires were then brought out through a plastic-filled trough.

7. *Bumpless Transfer.* Transferring system from manual control to automatic control and vice versa is performed readily by the operator with no rebalancing operation of any type. The control system remains continuously in balance so that any transfer is "bumpless." Should transfer from manual to automatic control be made with the process away from the control point, reset action brings the process rapidly and smoothly to control point with no proportional action bump. (The detailed circuitry showing how this is accomplished is not covered in this paper.)

References

1. THE EXPERIMENTAL BOILING WATER REACTOR. *TID-7012*, U. S. Atomic Energy Commission, Washington, D. C., 1957.

2. INSTRUMENTATION AND CONTROL OF THE EBWR POWER PLANT, W. C. Lipinski, R. L. Ramp. *Proceedings*, 11th Annual Instrument Conference, Instrument Society of America, Pittsburgh, Pa., Sept. 19, 1956.

3. AN ELECTRONIC CONTROL SYSTEM FOR INDUSTRIAL CONTROL APPLICATIONS, E. T. Davis. *Instruments*, Pittsburgh, Pa., June 1951.

Main Coolant Instrumentation for High-Performance Pressurized Water Reactor Power Plants

L. C. NOYES
MEMBER AIEE

Synopsis: New pressurized water reactor designs are obtaining higher efficiencies and smaller physical size by operating primary coolant loops at higher temperatures and pressures. Since the ultimate temperatures and pressures are limited by the critical point of water and instrument errors, highly accurate process-type instrumentation is required and the instrument engineer becomes an important member of the reactor design team.

A brief résumé of the design requirements for pressurized water reactor temperature, pressure, flow, and liquid-level instrumentation is discussed. The goals to be met by the development are followed in logical order by a discussion of the means chosen to achieve them. Because the primary use of high-performance pressurized water reactors at present is for ship propulsion, the requirements for reliability and long life under vibration and high shock have forced the evolution of self-balancing magnetic-amplifier circuits for process-type measurements as well as new detector designs.

In conclusion, performance data are presented on the measurement set errors.

THERE are two reasons for better instrumentation on a nuclear power plant:

1. To obtain greater plant efficiency and higher power output for a given reactor plant design.

2. To reduce the over-all plant investment by reducing the size and complexity of other plant components such as reactor core, piping, pumps, steam drum, etc.

For ship propulsion reactors, highest plant efficiency and greatest power output from a given reactor plant design are the most important. Generally, a decrease in efficiency means a larger increase in the size and weight of components such as reactor core, coolant pipes, pumps, etc., to obtain the same power output. Conversely, a relatively small increase in over-all plant efficiency can mean significant savings of weight and space which can be devoted to the purpose for which the ship is being built, be it merchantman or warship.

The primary use of pressurized water reactor power plants to date has been for ship propulsion, and the coolant instrumentation described in this paper was developed especially for shipboard use. However, the functional requirements for instruments of both shipboard and land-based power plants are the same, and while the physical construction might change, the principles of operation are equally advantageous for land based use.

The design of a pressurized water reactor power plant is too highly complex to even outline in this paper and has been reported in the literature.[1–6] However, it should be noted that the plant is an interdependent system and the selection of primary and secondary operating temperatures and pressures is a compromise of many conflicting factors. To achieve highest thermal efficiency, most effective utilization of nuclear fuel, and lowest cost for steam plant equipment, secondary steam system pressure and temperature should be as high as possible.

The primary coolant temperature, pressure, and flow are determined by such factors as size and cost of core, pumping power required, and steam-generator size and cost. The cost of primary system components increases with pressure and 2,500 psi (pounds per square inch) represents about the highest practical design pressure at present. The nominal operating pressure for a 2,500-psi design pressure has been 2,000 psi. At 2,000 psi the saturation temperature of water is 636 F (degrees Fahrenheit). With more accurate temperature and pressure instrumentation of the type described here, safety margins required to protect for possible instrument errors are reduced and the operating pressure may be safely increased to 2,200 psi with an increase of saturation temperature to 650 F.

Thus a decrease in measurement errors means an increase in power output for a given plant design, and the instrument engineer becomes an important member of the reactor design team.

Functions

The measurement of nonnuclear quantities on the main coolant loops of a typical pressurized water power plant are shown in Fig. 1. These include numerous temperatures and pressures as well as flow in the loop piping and level in the pressurizer vessel. Each of these measurements requires a local indication of the measured variable at the main coolant instrumentation cabinet as well as a remote indication on the main reactor plant control benchboard and possibly other locations such as the main coolant charging station. In addition to direct indication of the quantity, many of the functions initiate alarms, slow rod drivedowns, or scram the reactor through signals to the

---

Paper 58-702, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Chicago, Ill., March 17–21, 1958. Manuscript submitted December 23, 1957; made available for printing March 10, 1958.

L. C. NOYES is with the General Electric Company, Lynn, Mass.

This paper presents the work of a number of people in the Instrument Department and the General Engineering Laboratory of the General Electric Company. Both their work and their help in preparing this paper are gratefully acknowledged.

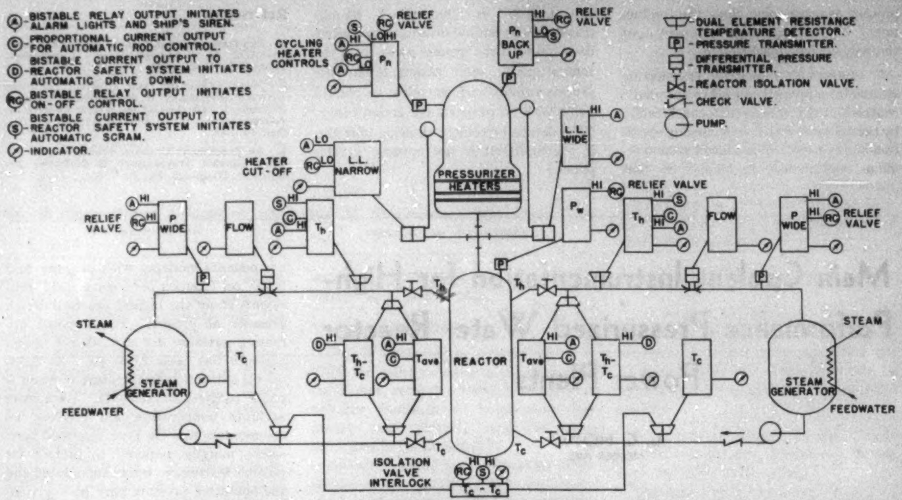

Ⓐ-BISTABLE RELAY OUTPUT INITIATES ALARM LIGHTS AND SHIP'S SIREN.
Ⓒ-PROPORTIONAL CURRENT OUTPUT FOR AUTOMATIC ROD CONTROL.
Ⓓ-BISTABLE CURRENT OUTPUT TO REACTOR SAFETY SYSTEM INITIATES AUTOMATIC DRIVE DOWN.
RC-BISTABLE RELAY OUTPUT INITIATES ON-OFF CONTROL.
Ⓢ-BISTABLE CURRENT OUTPUT TO REACTOR SAFETY SYSTEM INITIATES AUTOMATIC SCRAM.
-INDICATOR.
-DUAL ELEMENT RESISTANCE TEMPERATURE DETECTOR.
P -PRESSURE TRANSMITTER.
-DIFFERENTIAL PRESSURE TRANSMITTER.
-REACTOR ISOLATION VALVE.
-CHECK VALVE.
-PUMP.
RELIEF VALVE
CYCLING HEATER CONTROLS
Pn
Pn BACK UP
HEATER CUT-OFF
L.L. NARROW
L.L. WIDE
PRESSURIZER
HEATERS
Pw
RELIEF VALVE
P WIDE
FLOW
Th
HI
LO
STEAM
STEAM GENERATOR
FEEDWATER
Tc
Th-Tc
Tave
REACTOR
ISOLATION VALVE INTERLOCK
Tc - Tc

Fig. 1. Main coolant instrumentation for typical pressurized water power reactor

plant safety system when any of the quantities indicate that the measured quantity is out of control. Some quantities also operate either direct-acting on-off controls or provide a direct current proportional to the measured variable for rod control.

Design Requirements

The environmental and other design requirements for naval reactor instrumentation have been codified in *MIL-N-19900*.[6] A brief abstract of some of the more important points is included in Table I and this paper describes instrumentation built specifically to fill these requirements.

Vacuum tubes, motors, and most types of moving parts have been specifically prohibited except where absolutely essential in detectors and indicators since maximum reliability under these difficult operating conditions is required. The prohibition against moving parts and the requirement of highest possible accuracy under widely varying temperature, voltage, and frequency naturally lead to the selection of all-electric self-balancing systems using 100% negative-current feedback; see Fig. 2. Self-balancing systems of this type are current generators and provide all of the advantages of current output telemetering. This choice has been well justified since the systems have performed equal to or better than specification limits and far outperformed open-loop voltage output systems.

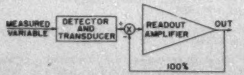

MEASURED VARIABLE
DETECTOR AND TRANSDUCER
READOUT AMPLIFIER
OUT
100%

Fig. 2. Self-balancing measurement system

In order to obtain maximum versatility and economy of manufacture, the measurement sets were split into four modular building blocks. These building blocks are:

1. Readout amplifier.
2. Bistable amplifier.
3. D-c constant-current source.
4. A-c regulator amplifier.

To aid in understanding the operation of the measurement sets, the operation of the four building blocks is described first, then the combination of the required blocks and the necessary detectors into a measurement set are described.

Since a number of proportional and on-off control functions may be required from a single measurement as well as local and remote indication, it was decided to standardize on a universal 0 to 5-ma (milliampere) d-c output from the readout amplifier into a maximum load of 5,000 ohms regardless of quantity being measured, or the range or span of measurement. This permits the use of one standard design of indicator and bistable control. It also provides sufficient power to operate as many indicators and on-off controls as required in addition to providing a current proportional to the measured quantity. In some cases as many as four indicators and six controls are operated from one measurement.

Readout Amplifier

For the readout amplifier a self-balancing magnetic amplifier was developed. It has two center-tap push-pull full-wave self-saturating stages in a 100% negative-current feedback-type (self-balancing) circuit. Magnetic amplifiers of this general type have recently been described

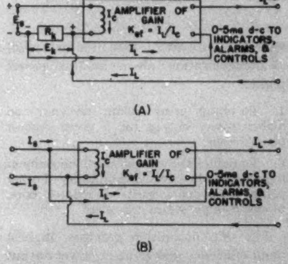

AMPLIFIER OF GAIN
$K_{ef} = I_L/I_c$
I_c
R_s
E_s
E_L
I_L
I_s
0-5ma d-c TO INDICATORS, ALARMS, & CONTROLS
(A)
(B)

Fig. 3. Self-balancing readout amplifier

A—$E_L \cong E_S$ B—$I_L \cong I_S$

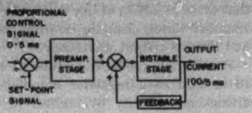


Fig. 4. Block diagram of bistable amplifier

by Geyger[7,8] and others.[9-11] A unique feature of this circuit is that the comparison between the output current I_L and the input signal always takes place ahead of the first magnetic-amplifier stage and only the difference signal I_c is fed to the input-stage control winding. (This puts any change in control-winding resistance or first-stage core characteristics inside the feedback loop, and zero drift is held to a minimum.)

The circuit offers several advantages:

1. When voltage comparison is used, as in Fig. 3(A), the negative-feedback voltage $E_K = I_L R_K$ is in series opposition to the signal voltage E_s, thus increasing the effective input impedance of the amplifier 200 to 300 times. Thus the readout amplifier presents a high-impedance load to the signal source and acts like an ideal voltmeter.

2. When current comparison is used, as in Fig. 3(B), the negative-feedback current I_L is in parallel opposition to the signal current I_S, reducing the effective input resistance of the amplifier 200 to 300 times. Now the readout amplifier presents a low-impedance load to the signal source and acts like an ideal galvanometer.

3. The effective internal impedance of the output circuit is much higher than its original value without feedback and the amplifier is essentially insensitive to changes in load resistance.

4. The output current I_L remains proportional to the input control current I_c and the amplifier is substantially insensitive to wide variations in temperature, voltage, and frequency.

For satisfactory closed-loop operation the time constant of the first stage is made as long as possible and the time constant of the second stage as short as possible. This has resulted in an amplifier having high open-loop gain, low zero drift, and complete freedom from instability under all environmental conditions.

Moderate precautions in matching magnetic core and silicon rectifier characteristics have resulted in a rugged amplifier capable of operating over the entire range of ambient conditions given in Table I with a zero drift referred to the input of less than $\pm 10^{-9}$ watts and an over-all accuracy of ±0.5% of full scale.

Standardizing on this readout amplifier design has permitted the elimination of separate voltage-regulated power supplies for the instrument cabinets and gives maximum economy of manufacture.

Bistable Amplifier

The alarm and on-off control functions provide a special problem. Here it is desired to initiate the alarm or control action when the measured variable reaches a predetermined value. Since the proportional control current output from the readout amplifier had been standardized at 0 to 5 ma d-c, it was possible to develop a bistable amplifier with a single input current range.

The bistable amplifier is an all-magnetic static switching element[12] designed to give a step change in output signal when the input signal reaches an accurate predetermined value. For use with the reactor safety system, it delivers a nominal output of 100 ma d-c into a 265-ohm load while the variable is in the "safe" control range. When the prescribed limit is reached the bistable stage "trips down" to less than 5 ma signaling the change to the reactor safety system. To ensure fail-safe operation, no relay is used because of the danger of relay malfunction under high shock or vibration, and the safety system is always energized when the variable is in the safe control range. Thus a power failure or open circuit in the bistable unit will initiate the protective action.

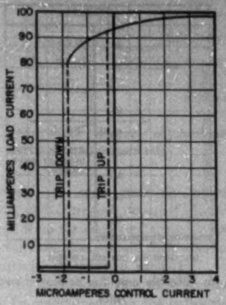


Fig. 5. Bistable-stage control characteristics

Table I. Environmental Design Requirements

Environmental Conditions	Reactor Compartment	Control Room
Radiation		
Neutrons	5×10⁷ neutrons/cubic centimeter/second maximum	negligible
Gamma rays	500 roentgens/hour, maximum	negligible
Temperature		
Air ambient	5 to 70 degrees centigrade	5 to 50 degrees centigrade
Coolant water	650 F maximum	
Humidity	100% maximum continuous	100% maximum continuous
Pressure		
Air ambient	10 to 30 psia	10 to 30 psia
Coolant water	2,500 psia maximum	
Roll and pitch	60 degrees from vertical, maximum; 10-second period	60 degrees from vertical, maximum; 10-second period
Vibration*	5 to 33 cps†	5 to 33 cps
Shock‡	2,000 foot-pounds, each axis	2,000 foot-pounds, each axis
Voltage		
Steady-state	115 volts ±10%	115 volts ±10%
Transient	20% added for 2 seconds	20% added for 2 seconds
Frequency		
Steady-state	60 cps ±5%	60 cps ±5%
Transient	3% to maximum of 6% for 2 seconds	3% to maximum of 6% for 2 seconds

* MECHANICAL VIBRATIONS OF SHIPBOARD EQUIPMENT. *Military Standard MIL-STD-167* (ships), U. S. Navy, Washington, D. C.

† Cycles per second.

‡ SHOCKPROOF EQUIPMENT, CLASS HI. *Military Specification MIL-S-901* (ships), U. S. Navy, Washington, D. C.

For the less critical on-off control and alarm functions, a relay is permitted. The bistable amplifier is identical with that just described except that the output current energizes the relay coil instead of the safety system. The relay is always energized when the variable is in the control range. Thus a power failure to the bistable unit drops out the relay causing the alarm action or securing the control function and the amplifier is again fail-safe.

The bistable amplifier[13] is basically a 2-stage magnetic amplifier consisting of a center-tap push-pull linear amplistat[14] preamplifier and a center-tap single-ended bistable amplistat output stage; see Fig. 4. To give the output stage its bistable characteristic, shown in Fig. 5, external positive feedback has been employed. For highly accurate switching the positive feedback is adjusted until the difference between the trip-up and trip-down points is made as small as possible and both occur very close to zero.

The preamplifier continuously compares its input signal to a preadjusted set-point signal. When the sum of the two signal ampere-turns is approximately zero it

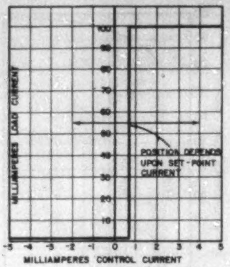


Fig. 6. Bistable amplifier control characteristics

drives the bistable stage through the trip point; see Fig. 6. The set-point bias is obtained from a Zener diode reference and is applied to the preamplifier in a direction opposing the readout amplifier proportional control current. By reversing the direction of the two control currents in the preamplifier and the direction of the signal from the preamplifier to the bistable stage, the bistable stage may be made to trip down on either an increasing or decreasing proportional control current from the readout amplifier. Thus the same bistable amplifier may be used for either a high or a low control and still maintain the proper fail-safe characteristics.

Since the portion of the preamplifier control characteristic shown in Fig. 7 represents only about 1% of the maximum proportional control current, the preamplifier is saturated most of the time. Therefore the gain of the preamplifier stage is high about the set point and the signal required to trip the amplifier is only about 4×10^{-8} watts.

Tests have shown the bistable amplifier to have a set-point accuracy of better than ±1% and a trip time of less than one second under the entire range of environmental conditions including high shock.

Where necessary, modifications of the amplifier gain and positive feedback may be made to provide variable or wide deadband characteristics. Also, additional input control windings may be provided for use where the coincidence of two or more signals is required to initiate a trip action.

D-C Constant-Current Source

The d-c constant-current source amplifier shown in Fig. 8 is a 2-stage self-balancing magnetic amplifier similar to the readout amplifier. It is used to supply a constant 400 ma into a maximum load of 52.5 ohms consisting of a temperature bridge and a feedback resistor. Since the bridge output voltage is directly proportional to the current applied to it as well as to the resistance thermometer temperature, it is imperative that the d-c constant-current source be of the same high-quality design as the readout amplifier.

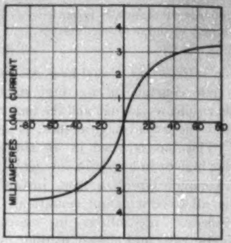


Fig. 7. Preamplifier-stage control characteristics

As shown in Fig. 8, a Zener diode is used as the reference signal source E_{ref} which is compared with the feedback voltage E_K. The difference of these two voltages is fed to the first stage as a control signal resulting in the control current I_c.

The output stage is a center-tap single-ended amplistat biased to operate at about 400-ma output for zero-signal input. Biasing the output stage in this way reduces the range of operation for the input stage, greatly simplifies the regulator operation, and takes maximum advantage of the high gain of the input stage around zero input.

Tests on amplifiers of this type have shown that any constant preset current can be held to a regulated accuracy of ±0.5% including the drift of the reference diode over the entire range of environmental conditions.

A-C Regulator Amplifier

The a-c regulator amplifier shown in Fig. 9 is similar in operation to the readout amplifier and the d-c constant-current source previously described, except that its output is a-c instead of d-c. It is used to maintain a constant current in the primary of the movable core transformers (MCT) of the flow and liquid-level measurement set detectors. The primary of MCT_1, with its core fixed in place, is connected in series with the amplifier load MCT_2. Its output voltage E_K is rectified and compared with a reference voltage E_s obtained from a Zener diode. This extra MCT_1 is necessary to correct for the waveform errors that occur in MCT_2 as line voltage to the amplifier varies, changing its firing angle and thus its output waveshape.

The d-c/a-c quotient amplifier in Fig. 10 is used as a regulator amplifier to divide one quantity by another in the following manner.

Let $I_s = X$, $I_K = YI_L$ and $I_L = K_{ef}(I_s - I_k)$; then

$$I_L = \frac{X}{\frac{1}{K_{ef}} + Y} \tag{1}$$

If K_{ef} is very large, then

$$I_L = \frac{X}{Y} \tag{2}$$

Over the entire range of environmental conditions, regulator amplifiers of this type have been proved to maintain a regulating accuracy of ±0.5% of full scale for variations of I_s from zero to full scale and for 4-to-1 variations in MCT_1 core displacement.

Temperature Measurement

The primary reasons for temperature measurement are to prevent overheating of the reactor core assembly or violent fluctuations in reactivity due to variations in water density changing the moderation. Overheating may occur in several ways:

1. Insufficient or loss of flow due to failure of pumps or closing of isolation valves.
2. Bulk boiling around the fuel elements reducing the heat transfer coefficient and causing unsafe pressures.
3. Excessive power demand from the steam generator causing fuel element internal temperatures to rise above a safe level.

The expansion of fuel elements and other components of the reactor core with overheating mayd istort or buckle the core radically changing the flow and flux distribution patterns. Hot spots then occur which rupture and burn out fuel element assemblies. Bulk boiling can generate large quantities of steam rapidly, causing the pressure to rise quickly to an unsafe value in the manner of a steam boiler explosion.

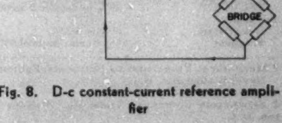


Fig. 8. D-c constant-current reference amplifier

Moderation of the neutron flux changes with the water density and reactor inlet water temperature must be monitored at all times to prevent cold water accidents. These occur if the reactor inlet water temperature is allowed to drop suddenly as might happen if isolation valves were suddenly opened to admit cold water from a loop that had been shut off for some time.

The temperature measurement sets are of three basic types:

1. A single-temperature measurement set (either of wide or narrow range) which gives an output proportional to a single temperature.
2. An average-temperature measurement set (generally narrow range) which gives an output proportional to the average of two temperatures.
3. A differential-temperature measurement set (generally narrow range) which gives an output proportional to the difference between two temperatures.

Single-temperature wide-range indication of reactor inlet temperature T_c is provided for start up. Also during start-up, indication of difference between pressurizer temperature T_p and reactor outlet temperature T_h is used to prevent bulk boiling.

Narrow-range measurement of reactor outlet temperature T_h provides indication, alarm, and scram for dangerously high reactor water temperature. A proportional control current from the T_h measurement set may also be used for automatic rod control.

Average temperature T_{ave} is used for indication and high alarm as well as to provide a proportional control current for rod control when the reactor plant is automatically controlled to a constant average temperature.

Loop differential temperature $T_h - T_c$ is a measure of the reactor thermal power output since flow is a constant value and the power delivered is:

$$\text{Btu/hour} = kF(T_h - T_c) \qquad (3)$$

where

k = a proportionality constant
F = flow, pounds/hour
$(T_h - T_c)$ = differential temperature, degrees

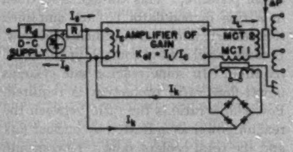


Fig. 9. A-c regulator amplifier

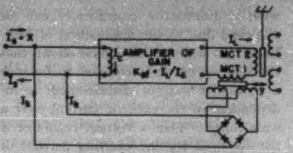


Fig. 10. D-c/a-c quotient amplifier

As previously mentioned, $T_p - T_h$ is used during start-up to indicate the difference between the reactor water temperature and boiling.

$T_c - T_c$ is used to interlock the loop isolation valves to prevent accidental opening of a cold loop and a cold water accident. In case of excessive difference in reactor inlet water temperatures it may also provide scram initiation signals to the reactor safety system.

The basic components of the temperature measurement sets consist of resistance temperature detectors (RTD), a resistance bridge circuit and its d-c constant-current power supply, a readout amplifier to amplify the bridge output and supply a direct current proportional to the measured quantity, indicators for local and remote indication, and bistable amplifiers to operate on-off controls and alarms arranged as shown in Fig. 11.

The RTD is a dual-element type with tungsten resistance elements and was especially developed for radioactive service.[15] Tungsten is used for the element in preference to platinum or nickel since, in a properly compensated and unbalanced bridge circuit with a constant-current input, it provides a linear signal directly proportional to the measured temperature. Its unique mechanical construction provides a high speed of response (0.5 second for 63%) for coolant flows above 5 feet per second as well as complete interchangeability between elements.

Use of a dual resistance element reduces the number of penetrations of the high-pressure piping to a minimum and allows either two temperature measurements at the same point or provides an installed spare.

The RTD in combination with its bridge, constant-current power supply, and readout amplifier provides a 0 to 5-ma d-c signal proportional to temperature from 0 to 700 F and of any span from full range to as low as 50 F. Measurements with a span of much less than 50 F are not practical since the fixed magnitude errors such as amplifier input-stage zero drift and RTD adjustment errors become dominant over the errors that are proportional to span such as indicator error, and increased accuracy does not result from the decreased span. Typical measurement ranges and their associated accuracies are shown in Table II.

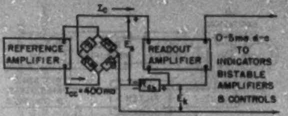


Fig. 11. Block diagram of temperature measurement set

Tests have shown that these accuracies can be maintained over the entire range of environmental conditions including high shock.

By proper choice of bridge balance point, the output current may be made to increase with either increasing or decreasing temperature for fail-safe operation. For a single-temperature measurement the RTD is located at R_1, for average temperature at R_1 and R_3, and for differential temperature at R_1 and R_4 in Fig. 11. The bridge transfer function is

$$E_0 = I_{in}\frac{R_2R_4 - R_1R_3}{R_1 + R_2 + R_3 + R_4} \qquad (4)$$

Pressure Measurement

The two types of pressure measurement sets used are the same except for the pressure transmitters which differ in range and span of measurement.

The wide-range pressure set (P_{wide}) measures loop or reactor pressure over the range of 0 to 2,800 psia (absolute). During a cold reactor start-up it is used by the operator while controlling the pressurizer heaters manually to bring the system up to normal operating pressure. In addition, it may be used to operate high-pressure alarms and electric relief valves if the system pressure rises above a safe value.

The narrow-range pressure set (P_{narrow}) measures pressurizer pressure over a range of approximately 1,800 to 2,500 psia. Its primary purpose is to provide on-off control for the pressurizer heaters to maintain the system within the normal range of operating pressures. In addition, it provides a more accurate indication of

Table II. Measurement Ranges and Associated Accuracies

Measurement	Range, F	Accuracy, F
T_c	0 to 600	±12
T_h	510 to 580	± 3.5
T_{ave}	495 to 545	± 3.5
$T_h - T_c$	0 to 50	± 4.5

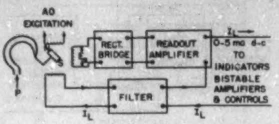


Fig. 12. Block diagram of pressure measurement set

system pressure as well as operating high- and low-pressure alarms, a low-pressure scram initiation, and an electric relief valve.

Frequently a P_{narrow} set is used to provide a safety backup measurement of reactor pressure in case of failure of the primary narrow-range pressure measurement. This narrow-range backup pressure measurement provides local indication of reactor pressure and separate channels for relief valve trip and low-pressure scram.

To provide a greater accuracy of pressure measurement, a new pressure transmitter has been developed for use in a self-balancing measurement system, shown in Fig. 12, with the readout amplifier previously described. In this unit a bourdon tube senses the pressure P and moves the magnet of a magnetic pickoff unit.

The magnetic pickoff unit is a saturable-reactor device having two active core legs with exciting control and signal windings surrounding each leg. When the excitation winding is connected to a source of alternating current, a voltage E_s is induced in the signal winding proportional to the difference between the steady-state magnetic fluxes through the two core legs. Similarly, the signal voltage falls to zero if the two fluxes become equal. Rotation of the permanent magnet from its null position by the bourdon tube produces an unbalance in the core fluxes which results in a proportional increase in the signal voltage. However, the fluxes may be rebalanced and the signal voltage again brought to zero by passing a direct current I_L of proper magnitude through the control winding. Thus the d-c control current necessary to restore the signal voltage to zero is proportional to the pressure applied to the bourdon tube.

Accuracy greater than that obtainable with a differential transformer can be achieved due to the application of feedback. In addition, tests have shown that the control current required to balance a particular magnet displacement is independent of the exciting current over a wide range. Thus a voltage regulator is not required for the exciting current of the magnetic pickoff unit as it is with a differential transformer.

Table III. Measurement Ranges and Associated Accuracies

Measurement	Range, PSIA	Accuracy, PSIA
P_{wide}	0 to 2,800	±76
P_{narrow}	1,900 to 2,500	±35

Adjustment of the output current I_L to the proper value is made automatically by the readout amplifier in a self-balancing circuit similar to that previously described for the temperature sets except that the balancing is done in the magnetic pickoff unit instead of across the resistor R_L, and a filter and rectifier have been added to eliminate interaction between the a-c signal voltage and the amplifier output-stage "hash." The small unbalance signal from the magnetic pickoff is amplified and delivered as a 0 to 5-ma d-c output current to the various indicators, alarms, and controls, and to the control winding of the magnetic pickoff.

The high current gain (K_{ef}) of the readout amplifier maintains the magnetic pickoff very close to balance at all times. Thus, the amplifier output current I_L is determined only by the position of the pickoff magnet and is practically independent of power supply voltage and frequency, amplifier linearity, and load resistance.

Typical pressure measurement set ranges and the associated accuracies for all environment conditions are given in Table III.

Flow Measurement

Flow is not necessarily a critical or extremely important measurement since the main coolant pumps generally can only be operated at half or full speed or shutoff. However, it is desirable to be able to detect and indicate loss of flow from unusual causes such as a loop isolation valve drifting closed or loss of a pump impellor.

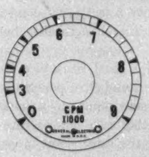


Fig. 13. Flow scale

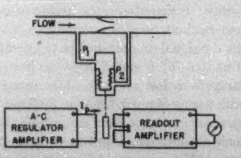


Fig. 14. Block diagram of flow measurement set

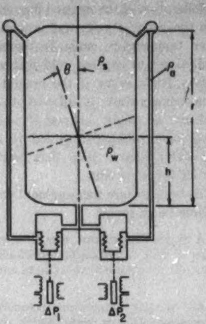


Fig. 15. Pressurizer piping schematic, wide-range external reference legs

In this case flow would be measured by detecting the differential-pressure drop across an elbow or other restriction in the loop piping. The differential pressure generated is:

$$\Delta P = kF^2 \tag{5}$$

where

F = flow, pounds/hour
k = factor which varies with water density and viscosity
ΔP = differential pressure, psi

However, since the accuracy required is not great (approximately ±10% to 20%) and the total flow in all loops is not required, a squared scale shown in Fig. 13 may be used satisfactorily, and density and viscosity compensation are not provided.

A block diagram of a flow measurement set of this type is shown in Fig. 14. In this set the differential pressure across the flow restriction is detected by the movement of a bellows to which is attached the core of an MCT. If the primary current in the MCT is held constant, as the core is displaced it will generate a voltage in the secondary proportional to the differential pressure. This voltage is rectified in a ring demodulator circuit, amplified if necessary, and used to indicate the flow. Usually, no alarms or control functions are required. In some reactor safe systems power-flow ratio protection is required. Power-flow ratio is the ratio between the reactor power as indicated by nuclear flux and the total coolant flow. To compute this ratio, a current proportional to total

Vol.
Pt. 1
195

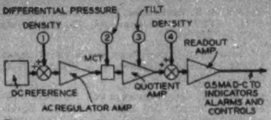


Fig. 16. Block diagram of liquid-level measurement set

flow is needed and this requires that the output current of each flow set be a linear function of flow. Linearization may be achieved where necessary by using compensating springs in the differential-pressure transmitter or by introducing nonlinearities in the readout amplifier.

Liquid-Level Measurement

Liquid level is not a very critical measurement; however, it is an important one and difficult to make accurately. Level is measured in the pressurizer vessel where it is desired to indicate the location of the interface between the water and steam. Indication of this interface is provided locally at the instrument panel, at the main coolant charging station where the operator manually operates valves to discharge or charge the system water as required, and at the main reactor plant control benchboard.

A high-level alarm is provided to indicate insufficient steam space in the pressurizer vessel. This is a hazardous condition since a sudden increase in system temperature would expand the water volume causing an insurge to the pressurizer. If this insurge were not relieved by compression of the steam, the system pressure would rise sharply and perhaps dangerously in the fashion of the familiar "water hammer."

Low-level alarm and pressurizer heater cutoff are also provided. During a severe down-temperature transient an outsurge of water from the pressurizer may uncover the top bank of heaters before the operator can charge the system with enough fresh water to keep the water level within its proper range. Under these conditions, the heater cutoff control turns off the heaters to prevent them from burning out while exposed.

Pressurizer level is usually measured by measuring the differential pressure between the bottom and top of the pressurizer vessel. A schematic arrangement of the pressurizer and its associated piping is shown in Fig. 15. External reference legs are provided with condensate pots which ensure a constant height of water column on the side of the differential-pressure transmitter. The other side is connected directly to the bottom of the pressurizer vessel. Generally, the differential-pressure transmitters are of the bellows type operating the core of an MCT and are similar to those used for flow measurement.

It can be shown for this configuration, assuming that the differential-pressure transmitters and all external piping lie in a plane that passes through the center line of the pressurizer and is perpendicular to the fore-and-aft center line of the boat, that

$$h=\left(\frac{\Delta P_1-\Delta P_2}{2}\right)\times\frac{1}{(\rho_w-\rho_s)}\times\frac{1}{\cos\theta\cos\phi}+r\left(\frac{\rho_a-\rho_s}{\rho_w\rho_s}\right) \quad (6)$$

where

P_1 = differential pressure, external leg 1
P_2 = differential pressure, external leg 2
ρ_a = density of water in external piping
ρ_s = density of steam in pressurizer
ρ_w = density of water in pressurizer
θ = angle of roll
ϕ = angle of pitch
h = height of interface from bottom of pressurizer
r = height of condensate pots from bottom of pressurizer

Thus it is seen that the desired height is a rather complex function involving the density of the water, the steam, and the roll and pitch of the boat.

A block diagram of the liquid-level measurement set which computes the desired height (h) is shown in Fig. 16. As is explained more fully in the following, the computation is carried out by first multiplying $(\Delta P_1-\Delta P_2)/2$ by the factor $1/(\rho_w-\rho_s)$, next dividing the result by the factor $(\cos\theta\cos\phi)$, and finally adding the correction term $r[(\rho_a-\rho_s)/(\rho_w-\rho_s)]$.

The actual differential pressure varies as a function of the pressurizer water and steam density. These densities may be computed from standard steam tables,[18] since the water is boiling, and are plotted as a function of system pressure in Fig. 17. At higher pressure and temperature, the steam density becomes large enough to be important. In fact, it may become as large as 25% of the water density in the operating range of pressures. This means that at some particular level where the height of the steam column is several times that of the water, the effect of the density variations will essentially cancel out and the level measurement will be accurate over wide variations in pressure. A typical pressurizer level density correction curve is shown in Fig. 18.

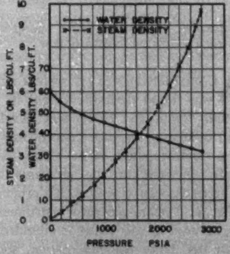


Fig. 17. Saturated water and steam density variation

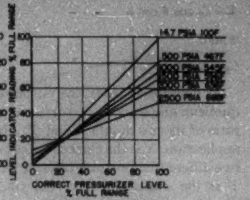


Fig. 18. Density correction curve for typical pressurizer level indicator

To obtain the first density correction, a signal is obtained from the narrow-range pressure measurement set over the normal operating pressure range. During start-up the correction is inserted manually by means of a potentiometer. Referring to Fig. 16 and equation 6, the density correction factor $1/(\rho_w-\rho_s)$ is added as a correction to the d-c reference signal and fed to the input of the a-c regulator amplifier. The output of the a-c regulator amplifier is then a function of density and is used to supply the primary windings of the MCT attached to the differential-pressure transmitters. Since their output is proportional to the product of primary current $1/(\rho_w-\rho_s)$ and core position $(\Delta P_1-\Delta P_2)/2$, the desired multiplication has been performed.

The signal from the differential-pressure transmitters is then fed into the quotient amplifier along with a signal from the tilt compensator. In a land-based plant $\cos\theta=\cos\phi=1$ and may be neglected. For shipboard use, a tilt compensator consisting of an MCT attached to a weight on a spring generates the tilt correction function

Table IV. Measurement Set Ranges and Accuracies

Measurement	Range, Inches of Water	Accuracy, Inches of Water
Wide range	0 to 70	±8
Narrow range	0 to 35	±3

$$Y = C_y \cos\theta \cos\phi \tag{7}$$

where

C_y = constant

As described earlier, the output of the quotient amplifier is proportional to the ratio of its two inputs X and Y. In this application, X is the current output of the two differential-pressure transmitters and Y is the displacement of the tilt compensator given in the foregoing.

After this division the quotient amplifier output is then fed to the readout amplifier input, where it is added to another density correction factor $r[(\rho_a - \rho_s)/(\rho_w - \rho_s)]$ which is generated by the d-c reference and the pressure signal. Thus the desired computations have been performed and the readout amplifier output is 0 to 5 ma d-c proportional to the liquid level h.

Due to the kind and number of manipulations required, liquid-level measurements are less accurate than those of temperature and pressure. Typical liquid-level measurement set ranges and accuracies under all conditions of tilt, pressure, and ambient conditions are shown in Table IV.

Conclusions

Improved main coolant instrumentation for a shipboard pressurized water reactor power plant has been developed using self-balancing amplifiers without moving parts. This instrumentation meets all the functional requirements and provides a high degree of reliability and long life under severe environmental conditions through the use of magnetic amplifier and silicon rectifiers.

Use of this type of instrumentation can reduce the reactor design safety margins required and allow more power output from a given design, thus providing significant savings in reactor cost, size, and weight.

The next step in reactor coolant instrumentation will probably be the application of transistors to this type of system in place of the present magnetic amplifiers. This change should result in improved performance, reduced size, cost, and weight, without sacrifice in the high degree of reliability obtained.

References

1. PROGRESS IN NUCLEAR ENERGY, SERIES II, REACTORS (book). "The Pressurized Water Reactor (PWR) Power Plant at Shippingport, Pa., Part A—Nuclear Power Generation," **J. W. Simpson, M. Shaw.** McGraw-Hill Book Company, Inc., New York, N. Y., 1956, pp. 289–315.

2. CONTROL OF NUCLEAR REACTORS AND POWER PLANTS (book), **M. A. Schultz.** McGraw-Hill Book Company, Inc., 1955.

3. THERMAL POWER FROM NUCLEAR REACTORS (book), **A. S. Thompson, O. E. Rodgers.** John Wiley & Sons, Inc., New York, N. Y., 1956.

4. DYNAMICS AND CONTROL OF NUCLEAR POWER PLANTS, **A. B. VanRennes, J. C. Simons, Jr., T. S. Gray.** *AIEE Transactions*, vol. 76, pt. I, July 1957, pp. 279–85.

5. SOME OPERATIONAL PROBLEMS OF THE NUCLEAR POWER PLANT, **A. R. Jones, R. L. Witzke.** *Ibid.*, Sept., pp. 373–77.

6. GENERAL REQUIREMENTS FOR NUCLEAR PROPULSION CONTROL AND INSTRUMENTATION EQUIPMENT. *Military Specification MIL-N-19900*, U. S. Navy, Washington, D. C

7. SELF-BALANCING MAGNETIC AMPLIFIERS OF THE DIFFERENTIAL-FEEDBACK TYPE, **William A. Geyger.** *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 39–46.

8. SIXTY-CYCLE LOW-LEVEL SELF-BALANCING MAGNETIC AMPLIFIER OF THE DIFFERENTIAL-FEEDBACK TYPE, **William A. Geyger.** "Proceedings of the Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 187–204.

9. A STABLE LOW-LEVEL MAGNETIC AMPLIFIER, **R. Cockrell.** *AIEE CP 56-737* (not available).

10. A DIRECT CURRENT TRANSFORMER WITH ADJUSTABLE RATIO, **H. E. Darling.** "Proceedings of the Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 236–44.

11. THE OPERATION OF THE CURRENT-TYPE SELF-BALANCING MAGNETIC AMPLIFIER, **Albert D. Krall.** *AIEE Transactions*, vol. 76, pt. I, Sept. 1957, pp. 380–84.

12. MAGNETIC AMPLIFIERS IN BISTABLE OPERATION, **L. A. Finzi, G. C. Feth.** *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 592–98.

13. A MAGNETIC AMPLIFIER LIMIT SWITCH, **D. L. Watrous.** "Proceedings of the Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Aug. 1957, pp. 256–264.

14. MAGNETIC AMPLIFIERS (book), **H. F. Storm.** John Wiley & Sons, Inc., 1955.

15. A RESISTANCE-TEMPERATURE DETECTOR FOR NUCLEAR REACTOR SERVICE, **F. R. Sias.** *AIEE Transactions*, vol. 76, pt. I, July 1957, pp. 363–65.

16. THERMODYNAMIC PROPERTIES OF STEAM (book), **J. H. Keenan, F. G. Keyes.** John Wiley & Sons, Inc., 1936.

Steam-Generator Instrumentation for a Nuclear Power Plant with Analog Computer Verification of Dynamic Action

E. E. LYNCH
FELLOW AIEE

D. P. WAITE
ASSOCIATE MEMBER AIEE

Synopsis: High-performance nuclear reactor designs of the pressurized-water type are used with smaller than conventional steam drums to convert from primary coolant heat to steam in a heat exchanger. "Swell" caused by steam bubbles changing in size and raising the liquid level various amounts at different steam-generation levels poses an acute feedwater control problem.

A solution to such a problem is discussed. The method of studying feasible solutions and the use of a computer setup, which eventually includes newly developed physical equipment replacing the initial simulated measurement components, is traced.

Performance data with actual electrical control initiation with various degrees of proportional and reset action are shown as it operates an electric valve operator to control a simulated feedwater flow.

A COMPREHENSIVE system study was found desirable to aid in the system design and component development of steam-generator instrumentation for a nuclear power plant. This paper illustrates some of the problems and solutions associated with such a program.

The term steam-generator instrumentation as used in this paper is intended to mean the major instrumentation, sometimes called level control or at times feedwater control, which is used to control the performance of the steam generator in supplying suitable quantity and quality steam to meet power requirements. With small steam drums used for space economy in nuclear submarines, the problem is not one of simple liquid-level regulation or of straightforward feedwater-flow control.

The problem has at least three facets. First, variations in level must be programmed. Second, changes in feedwater flow must not adversely upset the thermal equilibrium of the reactor coolant since the reactor power level depends on the moderating properties of

---

Paper 58-703, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Chicago, Ill., March 17–21, 1958. Manuscript submitted January 8, 1958; made available for printing March 10, 1958.

E. E. LYNCH and D. P. WAITE are with the General Electric Company, Lynn, Mass.

The authors wish to express special appreciation for the contributions made by their colleagues B. D. Leete and F. V. Reilley. Acknowledged also are the contributing efforts and co-operation of C. W. Taylor, B. F. Cassidy, and D. I. Welch of the Knolls Atomic Power Laboratory operated by the General Electric Company.

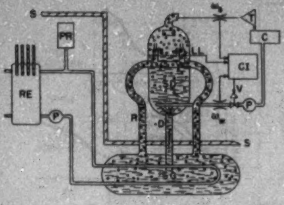

S
PR
RE
P
R
D
SD
S
LL
CI
C
T
V
P
ω_s
ω_w

Fig. 1. Simplified schematic of 3-element feedwater instrumentation: C=condenser, CI=control initiator, D=downcomer, P=pump, PR=pressurizer, R=riser, RE=reactor, S=shield, SD=steam generator or evaporator, T=turbine, V=feedwater valve, ω_s=steam flow, ω_w=feedwater flow

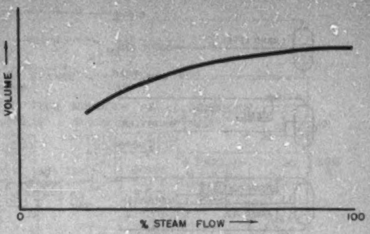


Fig. 2 (right). Typical swell curve of liquid volume in the steam drum versus per-cent steam flow

the coolant and the moderating properties are a function of the coolant temperature. Third, subcooled feedwater entering the steam drum as shown in Fig. 1 can alter the flow through the downcomer and its control is an important factor in determining optimum operation of feedwater control valves.

Instrumentation Problems

Plant Characteristics Influencing Instrumentation

Modern steam plant design emphasizes the severe plant conditions influencing steam-generator instrumentation and this is especially true of marine propulsion plants. For instance, the steam generator shown schematically in Fig. 1 may use a steam drum with a volume several times less than that in present electric power plants of comparable rating. Rapid power demand changes result in severe transients in thermodynamic conditions in the steam generator, and 3-element feedwater control is usually specified.

Naval service presents environmental conditions considerably more severe than those encountered in present normal conventional industrial plants. Rugged instrumentation is desired for both, however. Electric transmission of measurements and control with minimum moving parts and high reliability components with long trouble-free service must be available. Easy-to-use built-in test and calibration features are additional requirements.

The present trend of both mobile and stationary nuclear power plants for near-term operation is to continue to use the steam generator and turbine combination. This trend is compatible with boiling water, pressurized water, and other reactor designs, and the evolutionary growth of the steam plant has resulted in vast knowledge and experience with that portion of the plant.

Conventional steam plants tend toward constant temperature and pressure for all loads. However, reactor plants cannot always allow the associated rapid changes in coolant temperature which can accompany this type of operation. A constant average temperature of the coolant reduces such disturbing changes in moderating characteristics within the reactor and is one mode proposed for nuclear plants. The associated steam temperature and pressure available from the drum will therefore drop with load.

During a rapid decrease in steam flow with the accompanying supersaturation of the steam as pressure increases, there is possible increased moisture carry-over from the drum which may be harmful unless the equipment is properly designed.

The use of saturated steam (rather than superheated) introduces additional carry-over problems where dry steam may be essential for continued operation.

One of the problems associated with carry-over is that of swell. Swell or shrinkage is the result of steam bubbles expanding or contracting in the water in the heat exchanger and steam drum. It results in the change in position of the interface between water and steam. Fig. 2 shows how the water volume in a drum can change with steam flow due to swell. With small drums where the effect is most important, fast rates of change of steam generation are accompanied by nearly as rapid change in level due to swell.

Since the feedwater valve cannot drain water, increased level due to swell must be handled by the drum almost independently of level control. The control system can, however, tend to regulate the level to a steady-state condition where the level is in the optimum position to anticipate the most probable maneuver to be expected next from a liquid-level point of view. Thus, programming level as a function of steam flow is helpful.

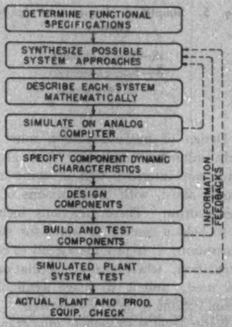

DETERMINE FUNCTIONAL SPECIFICATIONS
SYNTHESIZE POSSIBLE SYSTEM APPROACHES
DESCRIBE EACH SYSTEM MATHEMATICALLY
SIMULATE ON ANALOG COMPUTER
SPECIFY COMPONENT DYNAMIC CHARACTERISTICS
DESIGN COMPONENTS
BUILD AND TEST COMPONENTS
SIMULATED PLANT SYSTEM TEST
ACTUAL PLANT AND PROD. EQUIP. CHECK
INFORMATION FEEDBACKS

Fig. 3. Steps in typical approach to system design problem

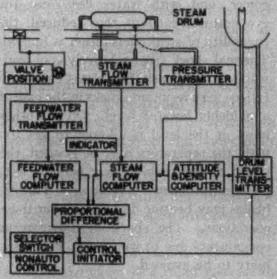

STEAM DRUM
VALVE POSITION
STEAM FLOW TRANSMITTER
PRESSURE TRANSMITTER
FEEDWATER FLOW TRANSMITTER
INDICATOR
FEEDWATER FLOW COMPUTER
STEAM FLOW COMPUTER
ATTITUDE & DENSITY COMPUTER
DRUM LEVEL TRANSMITTER
PROPORTIONAL DIFFERENCE
SELECTOR SWITCH
NONAUTO CONTROL
CONTROL INITIATOR

Fig. 4. Block diagram of 3-element feedwater instrumentation system

Swell is influenced by drum size. Its rate of change is influenced by the "internals" of the drum. Thus, variations of as great as four to one must be provided for. The rate of change of level with volume increase acts as a "gain" in the liquid-level feedback to the control and therefore may result in oscillation of the control unless designed to cover the range encountered. With mobile plants, change in this gain may result from list, trim, pitch, or roll, as well.

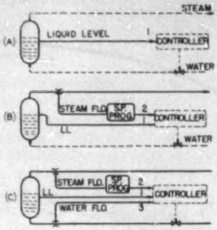


Fig. 5. Simplified diagrams of feedwater control

A—Single-element
B—Two-element
C—Three-element

Design of instrumentation must not only minimize oscillation but also it must not interfere, and in some cases should assist, with the functions of the steam drum. These functions include: 1. to hold a reserve of water for steam generation; 2. to provide a head to cause circulation, where forced circulation is not used; and 3. to make available a steam volume to help absorb swell.

Water level is extremely important to each of these functions. For instance, too low a level can lower the pressure head forcing water through the downcomer, and with water flow into the evaporator secondary thus reduced, the capability of steam generation is limited. Thermal shock from relatively cool feedwater on previously hot plates can cause fatigue with resultant cracks in the case of a suddenly lowered level. On the other hand, if the level is allowed to get too high, the separators may be flooded, the steam to the turbine may become wet, and "entrainment" may occur. Turbine blades may become damaged under these conditions. The shock of downstream components due to wet steam is a possibility to be avoided.

LEVEL CONTROL CONSIDERATIONS

To solve the problems posed by the foregoing plant characteristics, a system approach as indicated in Fig. 3 was followed. This was done since plant characteristics were expected to affect the dynamic action of the feedwater valve in a manner significantly different from that experienced in conventional plants; the instrumentation to be developed included sufficiently new approaches; and compatibility of plant and instrumentation characteristics was very essential to the necessary performance and reliability of the system.

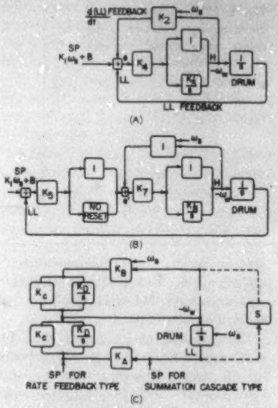


Fig. 6. Simplified transfer function diagrams of 3-element feedwater instrumentation systems

A—Rate-feedback type
B—Summation-cascade type
C—Equivalent diagram for both types

After determining functional specifications as dictated by plant conditions and steam requirements, a 3-element system approach was synthesized by plant engineers and is shown in block diagram in Fig. 4.

Mathematical equations which can be used to define such a diagram are

$$H=K_4K_3'\int(F-LL)_{dt}+K_4(F-LL) \tag{1}$$

$$F=K_2(\omega_s-\omega_w)+C \tag{2}$$

$$C=K_1\omega_s\pm B \tag{3}$$

where

H = valve opening resulting in proportional feedwater flow
LL = liquid level
ω_s = steam flow
ω_w = water flow
B = bias values to modify steady-state liquid level
K_1 = an adjustment to program the value of desired liquid level, as linear function of steam flow
K_2 = an adjustment which determines the relative magnitude of steam-water-flow mismatch feedback signal to liquid-level error signal
K_3' = an adjustment for reset rate of the control initiator
K_4 = an adjustment of proportional band

After simulating these in simple form on an analog computer and getting preliminary data on dynamic action, other possible system approaches were studied. Some considerations resulting from the study follow.

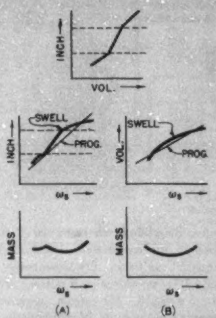


Fig. 7. Typical characteristic curves for 3-element feedwater instrumentation system

A—With level regulation
B—With volume regulation

Single-element, 2-element, and 3-element feedwater instrumentation are all available commercially to meet conventional land-based nonnuclear requirements as they presently exist.

Single-element control, although used satisfactorily with small boilers on some conventional installations, was found inadequate to meet requirements of the problem. Although extremely simple as shown in Fig. 5(A), the control depended on a steady-state level condition differing from the set-point level in order to sustain the feedwater valve at a given opening corresponding to a particular steam flow. With proportional band control only, this level condition was opposite that desired to accommodate swell with minimum change of mass in the steam drum. It was in the wrong direction to keep a high water level at high steam flows in order that as steam flow dropped, the level would not go down too much due to negative swell. It required high sensitivity tending toward excessive valve motion.

When reset action was added to bring the level back to a set point, a computer simulation of plant and instrumentation involved showed sluggish and slow dynamic movement of the feedwater valve when overpowering proportional action was used. With proportional plus reset action, bad oscillation occurred unless the action had such a long time constant that it was not useful in bringing level to its set value in a reasonably quick time.

Fig. 5(B) shows schematically a system which allows programming of the set point to accommodate swell with a low mass change in the steam drum. However, control problem experience indicates that a tachometer or rate feedback signal is

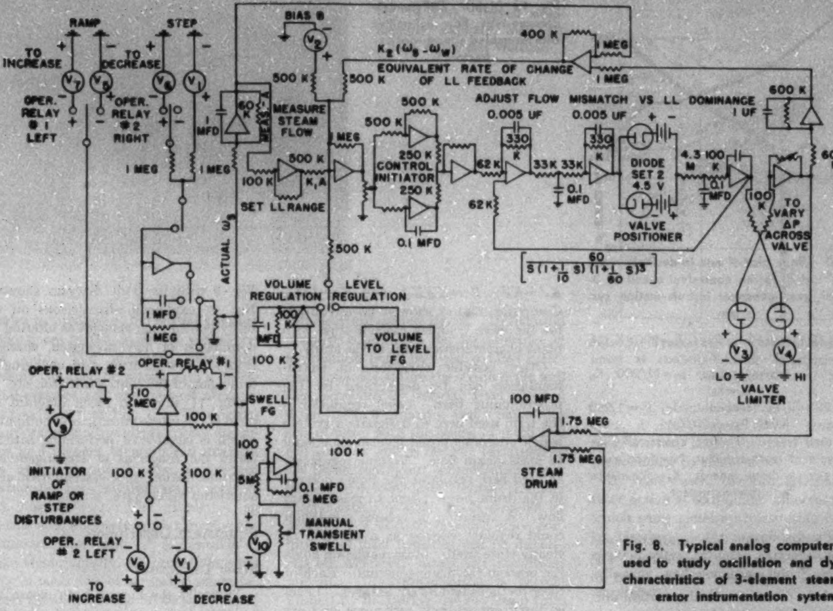

RAMP
STEP
TO INCREASE
TO DECREASE
OPER. RELAY #1 LEFT
OPER. RELAY #2 RIGHT
1 MEG
1 MFD
MEASURE STEAM FLOW
60 K
MEAS. A
BIAS B
V_2
$K_2(\omega_s - \omega_w)$
EQUIVALENT RATE OF CHANGE OF LL FEEDBACK
400 K
500 K
100 K
K, A
SET LL RANGE
CONTROL INITIATOR
250 K
0.1 MFD
ADJUST FLOW MISMATCH VS LL DOMINANCE
0.005 UF
330 K
62 K
33 K
DIODE SET 2 4.5 V
VALVE POSITIONER
4.3 M
600 K
1 UF
TO VARY ΔP ACROSS VALVE
$\left[\frac{60}{S(1+\frac{1}{10}S)(1+\frac{1}{60}S)^3}\right]$
VALVE LIMITER
LO
HI
V_3
V_4
ACTUAL ω_s
VOLUME REGULATION
LEVEL REGULATION
VOLUME TO LEVEL FG
SWELL FG
100 MFD
1.75 MEG
STEAM DRUM
OPER. RELAY #2
OPER. RELAY #1
10 MEG
INITIATOR OF RAMP OR STEP DISTURBANCES
OPER. RELAY #2 LEFT
5 MEG
MANUAL TRANSIENT SWELL
V_1
V_5
V_6
V_7
V_8
V_9
V_{10}

Fig. 8. Typical analog computer setup used to study oscillation and dynamic characteristics of 3-element steam-generator instrumentation system

usually necessary to provide dynamic performance that is fast enough to correct for large rapid disturbances and simultaneously operate within narrow limits from the set point. Two-element feedwater control did not inherently provide this signal feedback.

With three elements available for control as shown in Fig. 5(C), a rate feedback equivalent can be generated independent of the level transmitter signal. The difference between the steam mass flow and the water mass flow is used as a feedback. This is proportional to the rate of change of drum liquid mass. From the dynamic response standpoint, this feedback is better able to control the valve than is the level error signal. One reason for this is the fact that before the level error can control the valve, there must be a detectable error. This takes time. However, neglecting swell and programming of set point, that which is integrated with time to give a level error, i.e., rate of change of liquid level, is available immediately for control.

Another advantage of utilizing the rate signal with the level signal is the fact that it is one of the most effective ways known in servo techniques to counteract or compensate for an oscillating tendency of an otherwise single-element control, where reset is utilized in the control initiator to restore level to a set point or programmed value. This tendency results from the integrating effect of the steam drum. The drum and the reset component of the command signal result in two integrators in series (with 180-degree phase shift at all frequencies) and if reset rates are high, oscillation will result unless compensated for.

An advantage of obtaining the rate signal from steam- and water-flow measurements is the smoothness of the generated signal which is independent of random or cyclic rapid apparent level changes due to splashing, vibrations, bubbling, etc.

Fig. 6(A) shows simplified transfer functions of such a rate feedback 3-element control. K_4 represents a proportional band adjustment and K_3' an adjustment for reset action. The set point (SP) is programmed proportional (K_1) to steam flow (ω_s) with an adjustable level bias (B). H represents valve position and the simplification assumes this equal to water flow (ω_w). K_2 is an adjustment available to alter the dominance of the flow mismatch signal with respect to the liquid-level (LL) signal. To illustrate the need for a dominance adjustment, consider the effect of a rapid increase in power output calling for more steam. Swell will increase level almost immediately, calling for a more closed feed-

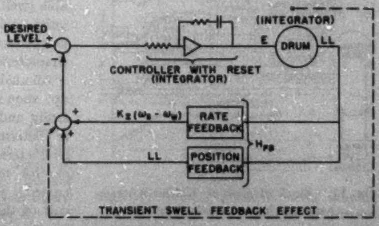

DESIRED LEVEL
+
−
(INTEGRATOR)
E
DRUM
LL
CONTROLLER WITH RESET (INTEGRATOR)
$K_2(\omega_s - \omega_w)$
RATE FEEDBACK
LL
POSITION FEEDBACK
H_{FB}
TRANSIENT SWELL FEEDBACK EFFECT

Fig. 9. Simplified diagram approximating closely the dynamic characteristics of rate-feedback type of 3-element steam-generator instrumentation system

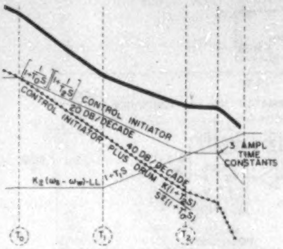


Fig. 10. Bode plot of gain in decibels versus T_0, ω_0, and P_0 for an equivalent circuit of 3-element steam-generator instrumentation system

$T_0=G_1/K_3'$ seconds/radian, $T_1=200\ K_2/N$, $T_2=1/K_3'$

$\omega_0=K_3'/G_1$ radian/second, $\omega_1=N/200\ K_2$, $\omega_2=K_3'$

$P_0=6.28\ G_1/K_3'$ seconds/cycle, $P_1=1{,}250\ K_2/N$, $P_2=6.28/K_3'$

K_3'=resets/second, T=time constants, G_1=finite gain of reset amplifier, N=drum gain/drum gain without innards, K_2=dominance adjustment constant

water valve. But flow mismatch will initially call for a more open valve. The dominance adjustment can be used to control the direction of valve motion during the initial stages of a maneuver.

Fig. 6(B) shows another approach to 3-element feedwater control. It is designated here as a summation cascade system since the liquid-level controller output is added to a flow mismatch signal to act as the input to the second controller. With reset in both controllers and with the integrating effect of the steam drum, 3-series integrators present a stability problem.

Fig. 6(C) is a circuit equivalent to Fig. 6(A) or 6(B) when K_A represents K_4 or K_5K_7; K_B represents K_2K_4 or K_7; $K_C=1.0$; and K_D represents K_3' or K_8'. Thus equivalence is established between this rate feedback and summation cascade control when the first cascaded controller has no reset action and when $K_7=K_2K_4$; $K_8'=K_3'$; $K_7=K_4/K_5$; and $K_5=1/K_2$. Computer studies showed no advantage for the more complex cascade type with plant characteristics used in the study.

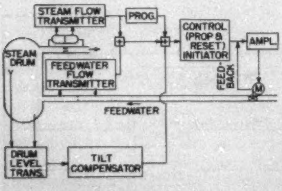


Fig. 11. Block diagram of 3-element steam-generator instrumentation system

Other possible variations in instrumentation can be determined by the programming used. Level regulation is the term used here to designate a system which programs liquid level proportional to mass steam flow. Volume regulation as used here regulates the liquid volume in the drum proportional to mass steam flow. A variation of volume regulation could regulate volume as a function of steady-state swell. Mass regulation holds the mass in the drum constant. Each of these methods had its advantages. Level is important to limit carry-over and exposure. Volume affects the available steam space to absorb swell. Mass establishes the hydrostatic pressure head that affects generator recirculation rates. Which regulation is best?

Volume regulation has a constant loop gain in the feedback from the drum detector. Level regulation results in a variable gain as shown by the variable slope of the inch-versus-volume curve of Fig. 7 which is caused by the internals of the drum. Level regulation is a practical form of control since detectors of level can be used linearly; computer studies indicate that stability is not a problem at the gains involved with plant characteristics usee in this study; and approximate mass regulation is possible by optimum linear programming as indicated by Fig. 7. To provide for optimum available swell volume, the volume would require nonlinear programming with steam flow which would depend on the swell characteristic.

An analog computer with a typical setup, shown in Fig. 8, was used in this study and proved to be a valuable tool in determining stability and dynamic characteristics of posssible variations in control techniques such as those discussed. In addition, the simplified circuit of the rate-feedback type shown in Fig. 9 with its Bode diagram shown in Fig. 10 explained observations on stability due to such changes as altering the dominance of flow mismatch, changing from low to high drum gain, changing the finite gain of the control initiator, etc.

Fig. 12 (right). Differential-pressure-type flow transmitter with assembled square-root extractor

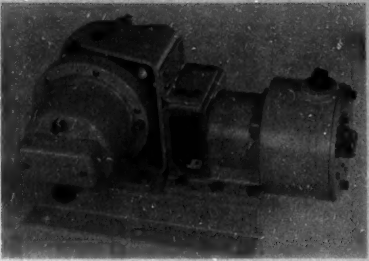


Fig. 11 shows in block diagram one form of the interconnections of equipment which is capable of performing satisfactorily the functions of steam-generator instrumentation of a system with characteristics which were studied.

Equipment Description

Flow Detectors

Typically the flow of both steam and boiler feedwater are measured using flow

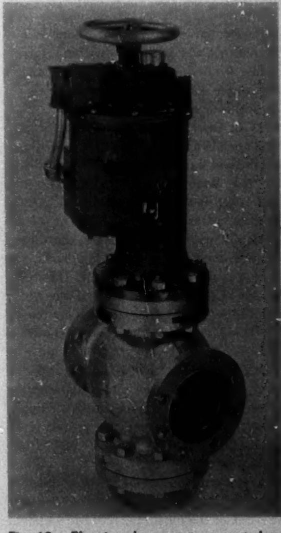


Fig. 13. Electric valve operator mounted on $3^1/_2$-inch regulator globe valve

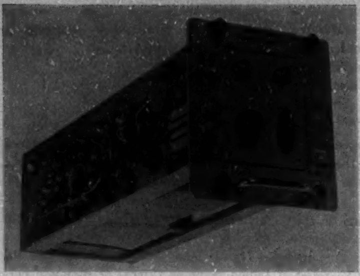


Fig. 14 (left). Typical measurement amplifier built for high shock and vibration

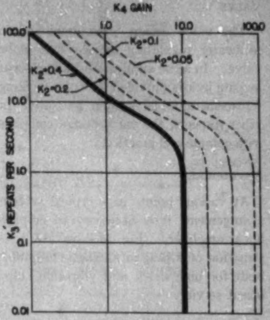


Fig. 15. Typical curves of locus of oscillation point as control initiator adjustments are changed

nozzles to produce a differential pressure proportional to the square of volumetric flow rate. The resultant differential pressure can then be sensed with a differential-pressure cell and pick-off arrangement as shown in Fig. 12. This design incorporates multigradient spring assemblies in a manner to provide a straight-line approximation to the square-root relationship between differential pressure and displacement. The result is an output voltage closely approximating the volumetric flow rate. It is desirable to make steam density compensations since error in mass flow of saturated steam becomes significant when no density corrections are made. To accomplish the correction, a bourdon-tube type of pressure detector provides an output voltage used for compensation. A straight-line approximation from varying spring gradients is also part of the pressure transmitter design. The advantage of this approach to square-root extraction as compared with other means is that simple noninteracting adjustments permit compensation for nonuniform elements elsewhere, such as flow coefficient calibration curves of the flow nozzles employed.

Liquid-Level Detectors

Liquid-level detectors often used may consist of a differential-pressure detector similar to that described for the flow detectors except that no additional spring clusters are employed to modify the linear relationship between differential pressure and electrical output signal. Differential pressure is produced by the difference in hydrostatic head between a filled reference water leg within the steam drum and that produced by varying level above a drum bottom pressure tap. Since this signal varies with the cosine of pitch and roll angle as well as density variations of the fluids, significant compensations may be desired in some installations. Density compensation is again made possible by measuring a system pressure since conditions of fluid densities can be directly related whenever saturation conditions prevail. Tilt compensation is accomplished by measuring the deviation from the vertical with a simple spring mass system and electric pick-off device. The output signal from the tilt compensator is divided into the signal from the differential-pressure cell in order to cancel the cosine factor in the signal.

Control Initiators

Equipment has been developed to produce proportional plus reset action using magnetic amplifiers. Adjustments include both proportional band (gain) and reset rate (coefficient of integration). The controller is actuated by an error signal that is also produced in this unit by comparing the steam-flow and water-flow signals to supply a rate feedback signal by adding this to a summation of measured liquid level. A manual level adjustment and a signal actuated by steam flow combine in a manner to produce a steady-state level command program with adjustable zero and range proportional to steam flow.

In some cases, the proportional plus reset action is considered unnecessary, in which case, necessary integrating is done by the servomotor associated with the electric valve actuator described in the following, and no proportional action is employed.

Valve Actuator

The mechanism of Fig. 13 is employed in a position-type loop when the proportional plus reset controller as described in the foregoing is used; a magnetic servo-amplifier provides excitation to the control phase of a servomotor in response to an error signal resulting from the comparison of position plus rate feedback of the controller output signal. Various valve actuator mechanisms have been designed to provide considerable flexibility in selecting the large power device associated with equipment described in the foregoing.

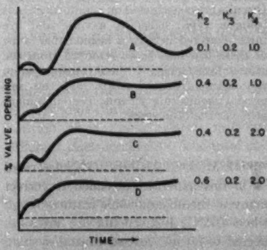


Fig. 17. Typical dynamic curves showing effect of varying control initiator adjustments

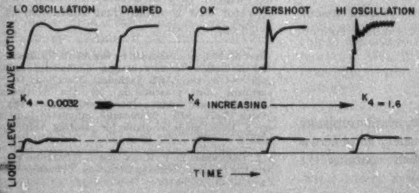


Fig. 16. Typical dynamic curves showing effect of changing control initiator gain adjustment, K_4

VALVE

The valve design often employed in feedwater regulating service is a globe valve. In some applications, a back-seat locking jack is provided to enable packing replacement under system pressure. In other respects, the valve design complies with commercial practice.

AMPLIFIERS

At various points in a typical system arrangement, it is necessary to employ linear amplifiers. Fig. 14 shows the construction of a plug-in magnetic amplifier built for high-shock and vibration shipboard service.

System Performance

During the system study, an analog computer was utilized to determine data on dynamic characteristics and oscillation. Fig. 8 shows one computer setup utilized to simulate a rate feedback approach. With this arrangement both step and ramp functions of steam flow were used as the disturbing characteristic. By utilizing a volume-to-inch converter or by short-circuiting it out, both volume and inch regulation as discussed were investigated and compared.

As physical instrumentation became available, the control initiator and valve simulation on the computer were replaced by actual equipment. Response characteristics taken under these conditions verified results taken with simulated equipment.

Typical of the data obtained were curves of dynamic motion of the valve and oscillation data for various values of equipment adjustments.

Fig. 15 shows typical curves of the locus of an oscillation point as adjustments are changed in a direction to increase the control initiator gain K_4, or its "repeats per second," K_3'. These are shown for various values of K_2 (the adjustment for changing the dominance of flow mismatch over liquid level in determining dynamic response).

Fig. 16 shows typical dynamic curves of valve motion with varying gain in the control initiator in response to a ramp increase in steam flow from 20% to 100% in a relatively short time and with level programmed approximately to equal the swell.

Fig. 17 shows typical dynamic curves of valve motion for various adjustments of the control initiator for a condition involving a large swell with a ramp increase in steam flow. It shows some of the variations obtainable by changing the dominance adjustment K_2.

Conclusions

This comprehensive system study of steam-generation instrumentation played an important part in the development of a system meeting the functional requirements dictated by the severe plant conditions which were imposed by nuclear submarines.

The dynamic and oscillation data obtained serve as typical of dynamic performance obtainable with nuclear power plant propulsion steam generators controlled with this form of instrumentation.

References

1. COMPUTER VERIFICATION OF STEAM GENERATOR INSTRUMENTATION FOR A NUCLEAR POWER PLANT, **D. P. Waite, E. E. Lynch.** "Proceedings of the Computers in Control Systems Conference," *AIEE Special Publication T-101*, May 1958, pp. 42–47.

2. ANALOG STUDY OF BOILER-REACTOR INTERACTION, **B. F. Borner, E. F. Cassidy.** *Nucleonics*, New York, N. Y., May 1957.

3. ALL ELECTRIC WATER LEVEL CONTROL FOR STEAM GENERATORS FOR NUCLEAR POWER PLANTS, **M. J. Kimball, S. F. Luna.** Instrument Society of America, Pittsburgh Pa., 1957.

Modular Concepts in Reactor Control Instrumentation

J. L. COCKRELL
MEMBER AIEE

D. F. RYAN
ASSOCIATE MEMBER AIEE

J. H. MAGEE
NONMEMBER AIEE

Synopsis: The basic circuitry of ORNL (Oak Ridge National Laboratory) reactor control instrumentation has been rearranged, and in some cases redesigned, so that more flexibility, reliability, and economy are achieved without sacrificing the safety performance of the original designs. Circuits are constructed on a uniform chassis (plug-in type) called a module and are interconnected in such a fashion that some of them can fail, or be removed from the system for checking, without causing system shutdown. In conjunction with an improved monitoring scheme, the equipment achieves the objectives.

THE MODULAR DESIGNS described in this paper are intended to correct many of the objectionable features found when ORNL instruments are interconnected to form a reactor control system. The four principal channels of this system are:

1. Safety system.
2. Pulse channel.
3. Logarithmic channel.
4. Linear or servochannel.

The basic amplifier circuits of ORNL instruments have been designed on modular chassis. This chassis is 6 inches wide and 12 inches deep and has an 8- by 8-inch front panel. A single 32-pin male connector on the rear of the chassis provides the necessary external connections. The module slides into a shelf assembly where it mates with a female connector.

Safety System

A typical 4-rod safety system utilizing four ORNL composite safety amplifiers is shown in Fig. 1. Each contains the following circuits:

1. Level preamplifier: a linear d-c amplifier which produces a voltage proportional to the ion chamber output current and hence reactor power level.

2. Period preamplifier: a linear d-c amplifier that develops a voltage inversely proportional to reactor period.

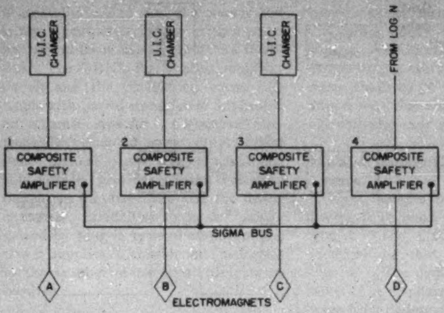
3. Sigma amplifier: a nonlinear d-c amplifier driven by either the level or period preamplifier. In Fig. 1, composite safety amplifier 4 has the period preamplifier driving the sigma amplifier. When the reactor power level is in the 0% to 100% region, the sigma amplifier output changes by 0.3 volt per 10% change in power level. For a 10% change in power level above 100%, the output changes by 0.6 volt.

4. Magnet amplifier: a linear d-c amplifier which controls the current to an electromagnet supporting a safety rod. This current varies inversely as the input signal from the sigma amplifier.

5. Fast-scram circuit: a linear d-c amplifier which operates a relay to warn the operator that a scram (fast shutdown) is being approached.

Paper 58-708, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Chicago, Ill., March 17–21, 1958. Manuscript submitted January 8, 1958; made available for printing April 7, 1958.

J. L. COCKRELL, D. F. RYAN, and J. H. MAGEE are with the Leeds & Northrup Company, Philadelphia, Pa.


U.I.C. CHAMBER
U.I.C. CHAMBER
U.I.C. CHAMBER
FROM LOG N
1
COMPOSITE SAFETY AMPLIFIER
2
COMPOSITE SAFETY AMPLIFIER
3
COMPOSITE SAFETY AMPLIFIER
4
COMPOSITE SAFETY AMPLIFIER
SIGMA BUS
A
B
C
D
ELECTROMAGNETS

6. Magnet power supply: an unregulated d-c power supply which provides the magnet current.
7. Amplifier power supply: a regulated d-c power supply which furnishes the plate and bias voltages for the amplifiers.

The only interconnection between the composite safety amplifiers is at the output of the sigma amplifiers. Here all four outputs are tied together creating what is commonly called a sigma bus. The safety amplifier with the highest output has control of the sigma bus. This connection provides a safety system mode of operation known as auctioneering. The one signal on the sigma bus is the input to all four magnet amplifiers and, therefore, every electromagnet current is determined by it. Consequently, a reactor scram is possible from any one safety amplifier. To ensure that the safety system is operating properly, a relay and meter measuring monitoring system has been designed into each composite safety amplifier. Some functions of the monitor warn the operator through an external annunciator if portions of the safety system have been disabled while others require a regular schedule of inspection through meter measurements.

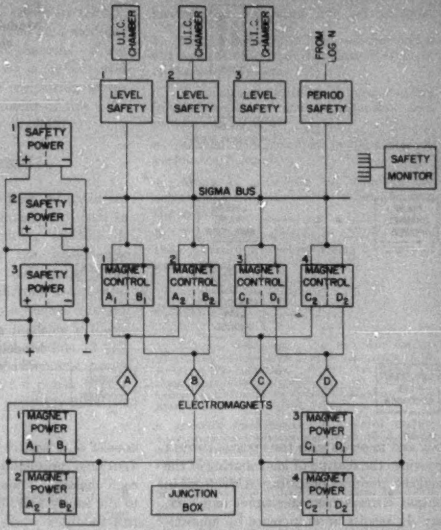
The principal objections to this safety system are:

1. An electronic failure in any of the composite safety amplifiers results either in automatic shutdown of the reactor or disables part of the safety system. In either case, unnecessary and annoying shutdowns result. For example, a burnout of the magnet power rectifier tube will automatically drop a safety rod causing reactor shutdown. Again, if the input tube to a magnet amplifier fails, a safety rod will not be able to be released except by the removal of magnet power (slow scram).

2. Protection against the consequence of certain sequences of component failures is not provided in the composite safety system. The composite safety system of Fig. 1 will "hang up" all four safety rods with the following sequence of component failures:

Fig. 1 (above). Four-rod safety system using composite safety amplifiers

Fig. 2. (right). Four-rod modular safety system


U.I.C. CHAMBER
U.I.C. CHAMBER
U.I.C. CHAMBER
FROM LOG N
1
LEVEL SAFETY
2
LEVEL SAFETY
3
LEVEL SAFETY
PERIOD SAFETY
SAFETY MONITOR
1
SAFETY POWER
2
SAFETY POWER
3
SAFETY POWER
+
−
SIGMA BUS
1
MAGNET CONTROL
A₁ B₁
2
MAGNET CONTROL
A₂ B₂
3
MAGNET CONTROL
C₁ D₁
4
MAGNET CONTROL
C₂ D₂
A
B
C
D
ELECTROMAGNETS
1
MAGNET POWER
A₁ B₁
2
MAGNET POWER
A₂ B₂
JUNCTION BOX
3
MAGNET POWER
C₁ D₁
4
MAGNET POWER
C₂ D₂

a. Loss of bias voltage in one magnet amplifier.

b. Open-plate load resistor in a second magnet amplifier.

c. Failure of input tube of a third magnet amplifier.

d. Failure of interstage tube in a fourth magnet amplifier.

Other combinations of failures exist which would have the same effect. The only annunciated failure is listed under item a.

3. Since each composite safety amplifier contains a level preamplifier, a period preamplifier, a sigma amplifier, and a fast-scram circuit, it can easily be shown that almost all safety systems using this equipment will result in many amplifiers serving no purpose at all. For example, a 6-rod safety system would necessitate six composite safety amplifiers. Normally three ion chambers would operate three of the level preamplifiers and three of the sigma amplifiers. One period preamplifier would derive its input from a single logarithmic channel and drive a sigma amplifier. Therefore, this system would have five period preamplifiers, three level preamplifiers, two sigma amplifiers, and two fast-scram circuits consuming power, but in no way a part of the system.

4. The physical construction of the composite safety amplifier is such that it is difficult to perform routine servicing and maintenance. Also, it cannot be disconnected from the system without causing reactor shutdown.

Modular Safety System

The proposed modular design of the 4-rod safety system as shown in Fig. 2 will correct these objectionable features.

First, the modules designed at level safety and period safety contain only a single preamplifier, a sigma amplifier, and a fast-scram circuit. Therefore, the number of safety amplifiers is exactly equal to the number of primary input signals. Each safety amplifier has its own filament transformer. Plate and bias voltages are obtained from the safety power modules with the positive voltage also providing the power for the UIC (uncompensated ion chambers). The signal outputs from the safety amplifiers are again auctioneered and, therefore, the voltage at the sigma bus is determined by the highest safety amplifier output. The conventional monitored information is fed to the junction box and external annunciation is provided should an abnormal condition arise. Because of the plug-in feature of the modules, it is then obvious that any level safety can fail or be removed for servicing without causing reactor shutdown. Directly below the sigma bus are the magnet control modules. Each module contains two identical magnet amplifiers whose outputs feed two different electromagnets. The total current in any one electromagnet is the sum of two currents from two magnet amplifiers located on different modules. For example, magnet amplifier output current A_1 plus magnet amplifier output current A_2 equals electromagnet current A. A feedback resistor located in the junction

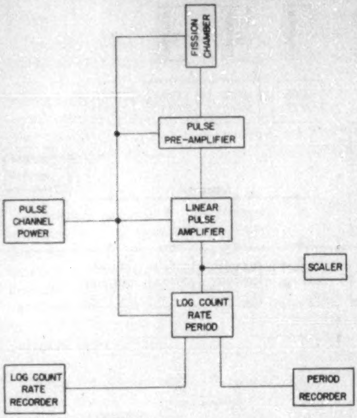


Fig. 3 (left). Modular pulse channel

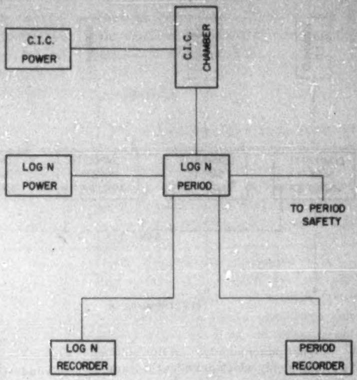


Fig. 4 (right). Modular logarithmic channel

box and in series with the magnet current provides the feature of load sharing of the magnet amplifier outputs. A decrease in output current from a magnet amplifier, say A_1, results in an increase by approximately the same amount in output current from magnet amplifier A_2. Of course, if the decreased output current is the result of a signal from the sigma bus, then both amplifier outputs are reduced. By virtue of this load sharing it can be seen that magnet control modules 1 and 3, 1 and 4, 2 and 3, or 2 and 4 can fail or be removed from the system without causing reactor shutdown. In order to ensure that a magnet amplifier output current will not increase to such an extent that no amount of input signal from the sigma bus is capable of releasing a safety rod, miniature-type thyratrons have been inserted into each magnet amplifier that will disable the output tube should this unsafe condition arise. The component failures listed under item 2 will cause the thyratron associated with each amplifier to fire, resulting in essentially zero output current from each amplifier.

The magnet power modules are the source of the electromagnet current. Each module contains two identical conventional unregulated power supplies. The total current in any one electromagnet is load shared between two power supplies located on different modules, and each power supply has the capacity of supplying the entire load current. Continued reactor operation is therefore possible if magnet power modules 1 and 3, 1 and 4, 2 and 3, or 2 and 4 fail or are removed from the system.

The modules designated as safety power consist of a positive and a negative conventional unregulated power supply on each chassis. These two voltages are fed to the safety amplifiers and magnet control modules where they are regulated and, therefore, furnish the necessary plate and bias voltages for the amplifiers. The total current demand from each supply is load-shared between three modules. In the 4-rod system shown in Fig. 2, two are capable of supplying the total demand. The failure or removal from the system of one can therefore be tolerated.

A monitoring scheme for this safety system has been designed on a separate module called the safety monitor. The front panel contains the following indicators and controls:

1. Four miniature side indicator meters which indicate continuously the outputs of the safety amplifiers. Three of these meters indicate the power level seen by each neutron-sensitive chamber and the fourth displays the reactor period.
2. Two neon lamps which inform the operator that the load sharing of the magnet power and safety power modules is correct. If it is not correct, this fact is annunciated and the lamps are extinguished.
3. Two wafer-type selector switches. One selects the function to be measured and the other selects the circuit under observation. Together with a panel monitor meter the following quantities can be measured:
   a. The cathode currents of every magnet amplifier tube.
   b. The output voltage of each safety preamplifier.
   c. The total current in each electromagnet.
   d. The sigma bus voltage.
   e. The output currents of every magnet power supply.
   f. The output current of every safety power supply.

Therefore, this monitor allows the operator to calibrate each safety amplifier, to locate incipient failures in the magnet amplifiers, and to locate power supply failures immediately.

Finally, it is possible to construct systems different in concept from the one described in this section. This can be accomplished simply by the manner of the external interconnection of modules. However, each new system evolved by this process must be carefully examined for safety and reliability characteristics.

Modular Pulse Channel

Fig. 3 is a block diagram of the pulse channel. Since the pulse preamplifier

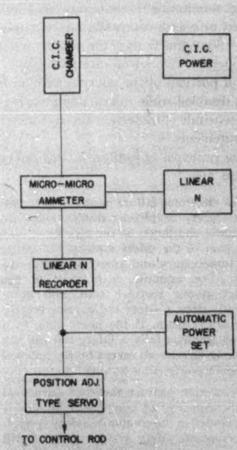


Fig. 5. Modular linear channel

is generally located near the reactor and is not a component of the control panel instrumentation, this unit was not modified. The ORNL linear pulse amplifier circuit and the ORNL log count rate meter with period circuit were duplicated on separate modules. A common electronically regulated power supply called pulse channel power was designed and this module provides all the necessary power for the amplifiers and the fission chamber. In addition to the advantages gained by plug-in type of amplifiers and power supplies, a considerable reduction in the number of electrical components resulted.

Modular Logarithmic Channel

The block diagram for the modular logarithmic channel is shown in Fig. 4. The ORNL log n pile period circuit was constructed on a module called log n period. An electronically regulated power supply called log n power was designed and this module supplies the necessary power for the log n circuitry. The compensated ion chamber (CIC) receives power from a modified ORNL CIC power supply designed on a modular chassis. Continuous adjustment of the compensating voltage by a helipot and range changing of the measuring meter affords very fine adjustment of this voltage and also easy reproducibility.

Modular Linear Channel

Three modules have been designed for the modular linear channel as shown in Fig. 5. The CIC power module is identical to that in the logarithmic channel. The display and control functions of a standard micromicroammeter have been designed on a second module. Remote location of the amplifier is therefore possible. The final module in this control system is the automatic power set. It is a regulated d-c power supply and constitutes the set point for the servo system.

Conclusions

The modular designs described in this paper provide a reactor control system having the following five advantages:

1. Many electronic failures will not result in reactor shutdown or disablement of parts of the safety system.
2. Sequences of electronic failures causing unsafe operation are considerably less probable to occur.
3. All electronic circuitry in the system is performing a function.
4. Many circuits can be removed for servicing and replacement without affecting the system.
5. The modular designs of circuits affords the opportunity of constructing a variety of systems.

References

1. TESTING PROCEDURES FOR REACTOR INSTRUMENTATION—COMPOSITE SAFETY AMPLIFIER, J. B. Ruble, S. H. Hannaher. *Report*, Oak Ridge National Laboratory, Central Files no. *CF56-5-30*, Office of Technical Services, Washington, D. C., Mar. 7, 1955, sect. P, pp. 1–13.

2. TESTING PROCEDURES FOR REACTOR INSTRUMENTATION—A-1 LINEAR AMPLIFIER, H. E. Banta, S. H. Hannaher. *Ibid.*, Feb. 8, 1954, sect. A, pp. 1–17.

3. PROCEDURE FOR TESTING LOG N AMPLIFIER, H. E. Banta. *Q-915*, Oak Ridge National Laboratory, Office of Technical Services, Washington D. C.

Variation of the Trip Point in the ORNL*-Type Safety System

J. L. COCKRELL
MEMBER AIEE

C. W. RICKER
NONMEMBER AIEE

IN nuclear reactor safety systems electromagnets are often used to connect neutron-absorbing rods to positioning drive units. The currents through these electromagnets can be varied in a continuous or discontinuous manner as a function of the reactor neutron level, the ultimate objective being to enable a trip or scram should an unsafe operating condition be present or imminent. A group of instrument channels, interconnected so as to determine whether the "unsafe operating condition" exists and, upon such determination, to trip the rods (all with no operator attention), comprise a safety system. One such safety system is the so-called auctioneering array of safety amplifiers, all deriving information from separate channels, and jointly feeding a single common point of interconnection or bus. The state of this bus is determined by the highest bidder (the channel with the highest output voltage). Thus a single channel, detecting an unsafe condition, is able to assume control and eventually shut down the system.

The functioning of the individual channels, their inputs, and the interaction of a group of these channels, must be known so that due consideration may be given in setting up the equipment to cause trip to occur under the proper conditions. It is the objective of this paper to supplement the understanding of single-channel performance by presenting the significant aspects of system performance.

Safety Channel

The composite safety amplifier, shown in Fig. 1 (ORNL *Q-1565*,) is representative of a class of instruments designed to perform the safety functon described hereinbefore. It contains both a level preamplifier and a period preamplifier, but one of these two is used at a time. The composite safety amplifier is a single channel in itself and is generally used in a system of three or four channels which are shown in block diagram form in Fig. 2.

The right-hand composite safety amplifier has the period preamplifier connected; the other amplifiers have their level preamplifiers in use.

The preamplifiers, whether level or period types, are linear and are designed to accept input signals proportional to the neutron flux of the reactor. The level preamplifiers receive currents up to about 50 microamperes at 150% full power from ion chambers (suitable for about three decades below full power), and deliver output voltages of 22 volts at zero power and $22+15$ or 37 volts at 150% full power. The period preamplifier derives its information from the log n amplifier (suitable for about six decades below full power). This input, in the form of a voltage, is first differentiated and, after amplification, is delivered as an inverse period signal which covers exactly the same range of voltages as delivered by the level preamplifiers. Here an infinite

---

* Oak Ridge National Laboratory.

Paper 58-709, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Chicago, Ill., March 17–21, 1958. Manuscript submitted December 23, 1957; made available for printing March 24, 1958.

J. L. COCKRELL is with the Leeds & Northrup Company, Philadelphia, Pa., and C. W. RICKER is with the University of Michigan, Ann Arbor, Mich.

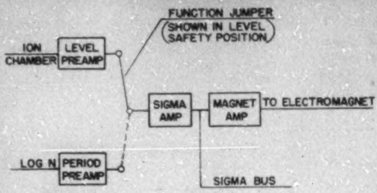


Fig. 1 (left). Composite safety amplifier (signal circuits)

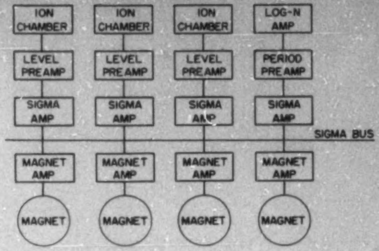


Fig. 2 (right). ORNL-type safety channels and system

period, corresponding to zero rate of change of reactor power, will be evidenced by 22 volts at the period preamplifier output. A period of about one second will cause a preamplifier output of about 37 volts, depending on the exact differentiating circuit elements and the amplifier gain.

The sigma amplifiers are identical and establish the voltage of the point of system interconnection—here called the sigma bus. This voltage is the input to all the magnet amplifiers, which are linear and in turn supply currents to the individual electromagnets. The sigma amplifier herein described is of particular interest and is considered in detail, both as a component and as an interconnected part of the over-all safety system.

Even when the magnet currents are continuous functions of reactor power level, as in the composite safety amplifier system, it is desirable that they vary only a small fraction of their nominal zero power value when the reactor is experiencing normal power excursions or transients. In order that the magnet currents will decrease rapidly when abnormal power excursions or transients occur, a nonlinear link in the channel is necessary. The safety or sigma amplifier is this nonlinear link, and it carries the dual role of generating the input-output nonlinear relation and of providing coupling between the individual channels. All other parts of the circuit, before and after this sigma amplifier, are linear and do not interact between channels except, of course, through the reactor power. Fig. 3 shows the typical circuit arrangement of the sigma amplifier, and Fig. 4 is the approximate characteristic of a single such amplifier.

V_1 is a conventional cathode-follower driver, which delivers the input voltage to the sigma amplifier tube V_2 through R_4. V_3 is a clamping diode which maintains the grid voltage and thus the output voltage essentially constant for low signal inputs. This amplifier is designed to be operated with the level preamplifier and ion chamber so that each input volt above the 22 volts shown is equivalent to 10% of reactor power. Thus, full power occurs at 32 volts and the overload power of 150% occurs at 37 volts. Note that the diode clamping action ceases at about 70% or 80% of full power. When driven by the period preamplifier an input of 22 volts corresponds to an infinite period while 37 volts corresponds to the trip period (usually one second).

Conventionally, the sigma amplifier is followed by a linear magnet amplifier which decreases in output electromagnet current about 5 milliamperes for every volt rise in the sigma bus. The actual magnitude of magnet current, for any input voltage, can be adjusted, and is generally set to give a trip or scram at a particular value of voltage input to the magnet amplifier. This voltage input is the same as the output voltage of the sigma amplifier and is the point of system interconnection. Thus, the name sigma bus has been attached to this point.

Interconnection and Interaction

The output impedance of a cathode follower is normally expressed in terms of the parameters of the circuit and tube, and is in the usual form

$$Z_{\text{out}}=\frac{R_K r_p/\mu}{R_K+r_p/\mu}=\frac{r_p}{\mu+\dfrac{r_p}{R_K}} \qquad (1)$$

where R_K is the value of the cathode resistor, r_p is the dynamic plate resistance of the tube, and μ is the amplification factor.

If r_p is not a constant, the impedance presented by the cathode follower to an incremental voltage change, impressed on the cathode of the tube, is more conveniently expressed as

$$Z_{\text{out}}=\frac{R_K}{\dfrac{\mu R_K}{r_p}+1} \qquad (2)$$

where r_p appears only once.

When the cathode is being driven positive by a voltage from an external source (for instance, another cathode follower), the tube current decreases, r_p increases, and the Z_{out} value approaches the value of R_K. On the other hand, when the cathode is driven negative, the tube conducts more strongly, r_p becomes stabilized, and the Z_{out} approaches r_p/μ (for the usual case when $\mu R_K/r_p>>1$). This variation is shown in Fig. 5 for a typical *Q-1565* sigma amplifier.

The significance of this change in output impedance, when several of these amplifiers are connected together at the cathodes, is that some interesting system results are possible. In particular, if a single tube cathode voltage increases because of a positive-going voltage on its grid, this action tends to decrease the impedance of the driving tube and at the same time to increase the loading impedance of the driven tubes (decreasing the loading). One might ask if (in the limit) the loading of the driving tube by the driven tubes could be made negligible. The data for Fig. 6 were taken in the laboratory on a simulated system of four sigma amplifiers and driving stages in order to establish a quantitative basis for discussing these notions. These curves for various loading conditions show that although the condition of no loading is approximated under some conditions, significant loading does occur, and that (if accurate knowledge of the interconnected cathode voltage is desired) this loading must be carefully considered.

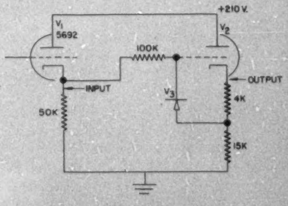


Fig. 3. Sigma amplifier

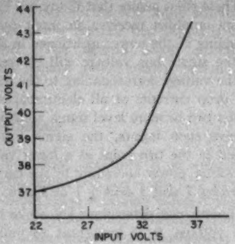


Fig. 4 (left). Sigma amplifier characteristic

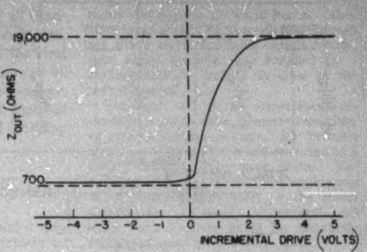


Fig. 5 (right). Input impedance variation

For the case of this type of circuitry in a nuclear reactor safety system, the final result is variation in the trip or scram point, depending on the nature or combination of inputs to the contributing safety amplifiers.

Normal Extremes

The safety system of a typical swimming-pool research reactor might contain two or three level safety channels and one period channel (as shown in Fig. 2). This gives rise to two normal extremes of safety action. One is the case arising from a very slow rise to an overload condition, but where the period safety amplifier is supplying negligible information. The level channels will be receiving the information from their separate ion chambers corresponding to this rising power level. If the channels are properly adjusted and calibrated, they will show a simultaneous slow rise in cathode voltage. (The level channels are driving the sigma bus upward in voltage.) The period safety, on the other hand, is receiving no significant information, and its cathode will be driven by the rising cathodes of the level channels. Any loading that would occur is by the single period channel alone (or all period channels if there should be more than one). This loading for the instances of two or three level channels is shown in Fig. 6 by the curves titled *2 driving 1* or *3 driving 1*. In either event the departure from the *unloaded* curve (single-unit characteristic) is quite small.

The other extreme is that encountered when the reactor power level is a decade or more below full power, but is rising rapidly enough to cause significant output from the period safety channel. Here the single period safety amplifier (or all such, if there are more than one) would act as the driving channel, while all level channels would be driven through the cathode interconnection; see curves labeled *1 driving 2* or *1 driving 3*, Fig. 6. In this instance the loading impedance seen by the driving period amplifier would be the impedance of all other amplifiers connected in parallel.

These two extremes for the example of a system using three level and one period channels are shown in Fig. 7(A). Also shown against the sigma bus voltage level is the linear magnet amplifier characteristic, Fig. 7(B). A rising sigma bus voltage corresponds to a falling magnet current. It is important to note that, for a given magnet current adjustment, the point at which the electromagnets release is determined only by the sigma bus voltage level.

Figs. 7(A) and (B) are read for a given situation (for example, *3 driving 1*) by noting a given input signal V_1 produces a sigma bus level L_1 which causes a magnet current I_1. A higher input signal V_2 produces a sigma bus level L_2 and a magnet current I_2. Note again that when the magnet current adjustment has been made, the position of the magnet amplifier characteristic in Fig. 7(B) is deter-

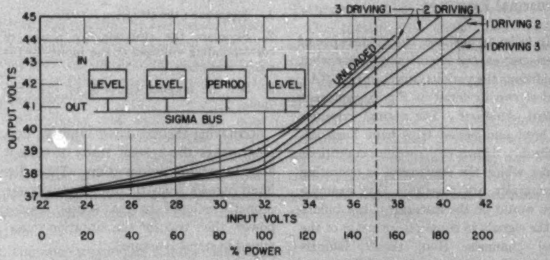


Fig. 6. Loading characteristics of safety amplifiers for various driving conditions

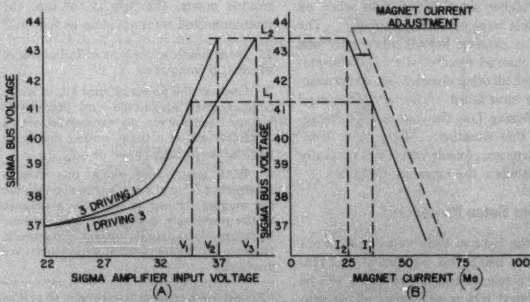


Fig. 7. Over-all transfer characteristic

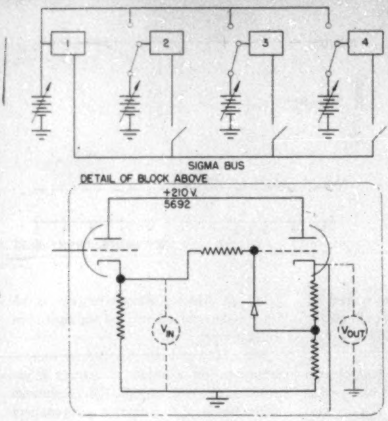


Fig. 8. Interconnection test scheme

mined and the sigma bus level must be up to L_3 to cause a trip. If only one safety amplifier receives an input (*1 driving 3*), its input must proceed to V_3 before the trip point is reached.

Abnormal Extremes

In order to evaluate the significance of these curves, one must know under what conditions the various curves in Fig. 7(A), and shown in detail in Fig. 6, represent actual situations. The normal extremes of level and period trips have been described. However, certain conditions under which the equipment is operating abnormally are possible. One example here would be the instance of the failure of the signals to reach one or more of the level channels. Here those channels which did receive signals would have to drive those which did not, in order for the system to reach a trip. Under these conditions the trip point would be reached at a higher signal input than when all channels were operating normally. The obvious extreme here is when only one level channel receives the level information and all other channels are either inactive or have failed. The curve *1 driving 2* or *1 driving 3*, as the case may be, represents this situation. Many other combinations are possible which will generally fall between the extremes described.

System Setup Procedures

In the light of the foregoing, it is not possible, using this equipment, to prescribe a reactor trip power which will be the actual trip power under all circumstances. Accordingly, it is well to examine how the trip point is inserted into this safety system, what the alternatives for the setup might be, and the resultant actual trip points.

Qualitatively, setting the trip can be summarized by three distinct steps:

1. Determine the situation for trip.
2. Simulate the situation by means of corresponding voltages at the inputs to the sigma amplifiers.
3. Set the magnet currents to the drop value.

Removing the simulated input voltage now makes the system ready to accept actual operating information. When the input signals from the ion chambers are properly related to the actual reactor power, then trip for the situation determined in step 1 is assured.

A particular example of carrying out these steps is helpful in bringing to light some of the details involved. If the criterion for trip is an absolute maximum reactor power, this trip is set into the interconnected safety system as follows:*

1. Determine the per cent of full power at which trip is required.
2. Convert the power of step 1 to a simulated voltage magnitude and apply this voltage successively to each level sigma amplifier (one at a time), noting the sigma bus voltage in each instance.
3. With the lowest sigma bus voltage determined in step 2 applied to the bus, set the magnet current in each electromagnet to the previously determined drop current by means of the magnet current adjustment. (Be sure that all bus protect adjustments are disabled.)

These steps assure that if any one level sigma amplifier receives an input corresponding to the power mentioned in step 1, the sigma bus voltage will be raised to the voltage corresponding to the magnet drop currents of all electromagnets. When two or more level sigma amplifiers receive such inputs, the sigma bus is raised to the trip point at a lower value of reactor power according to the curves *2 driving 1* and *3 driving 1* of Fig. 6.

Having once determined the sigma bus trip voltage, it is important to understand that the reactor period which, operating through the log *n* circuit and the period safety circuit, raises the sigma bus to this trip voltage is the trip period. When the sigma bus trip voltage is reached by the period sigma amplifier acting alone (at low power), the reactor period causing this rise is the limiting period and must be consistent with the operating and safety philosophy at the site concerned. At higher powers, approaching full power, the level sigma amplifiers contribute to the sigma bus voltage—and a slightly longer period will trip the system. Furthermore, for a given sigma bus trip voltage and any specified loading condition (both as described in the foregoing), the period for trip can only be changed by altering the components in the differentiating circuit at the input to the period preamplifier or by changing the gain of that particular period preamplifier.

Some figures from the illustrations may point up these facts. Referring to Fig. 6, if the trip is set with a 150% power signal (37 volts) into a single level sigma amplifier, then the trip is assured at 150%, and the sigma bus voltage for this will be 41 volts. When three level safety channels were operating together (*3 driving 1*), the trip (sigma bus raised to 41 volts) would occur at 126% of full power. As an alternate safety setup procedure, it might be assumed that three level channels would be operating normally to bring the sigma bus voltage up during a power excursion. To simulate this condition, the 150% power signals would be applied at the same time to the inputs of all three level sigma amplifiers. Note that the sigma bus voltage for trip is now 43 volts. Now a power rise to 150% full power would trip the system when all channels are operating normally, but it would require a rise to 180% of full power in order to ensure the trip if for any reason only one sigma amplifier could see the power rise. If a 1-second period is required to raise the sigma bus to 41 volts, the same period circuitry would require a period of about 0.83 second to raise the bus to 43 volts. Thus for the first setup situation in this paragraph the trip period is one second. The second setup situation would require a 0.83-second period for trip.

---

* Normal bench test and description of all routine setup procedures are described in ORNL Central Files no. *56-5-30*, section P.

Conclusions

From these examples it is apparent that the actual trip power can vary over a significant range of reactor power. However, the highest power which the system could allow under any combination of circumstances (except failure of all channels) is determined by the situation present if only one channel receives the level information. Since this situation can be simulated during the setup procedure, the ultimate trip point can be specified and becomes an essential part of the setup procedure.

Appendix. Simulated System

Measurements of system interaction were made on a simulated system of four identical sigma amplifiers as shown in Fig. 8. Also, since the clamping currents will flow through the diode and the driving output circuit, each sigma amplifier was supplied by a cathode follower exactly as done in the actual system.

The input and output connections were made in ways to allow complete versatility of interconnection scheme. All voltages were read within 1/4 volt and corrected by a 0.25% Rawson Polyranger Meter. The readings on the sigma bus where loading was being observed were made on the Radio Corporation of America Voltohmyst. All other readings were made by accurately calibrated panel meters.

An Integrated Mobile Microwave Communications System for the Massachusetts Turnpike

RALPH F. LOWE
ASSOCIATE MEMBER AIEE

A COMMUNICATIONS system was conceived and designed which would meet the special requirements of modern toll road operations simply and within reasonable costs.

The Massachusetts Turnpike Authority had requirements for the following services:

1. Telephone service between all maintenance centers and administrative headquarters.
2. Telephone and teletype service between all police barracks and administrative headquarters.
3. Vhf (very-high-frequency) mobile communications:
    a. between cars
    b. between central dispatch and cars
    c. between central dispatch and interchanges
    d. between interchanges and cars
    e. between auxiliary dispatch points and cars or interchanges

An integrated vhf mobile-microwave system, which would meet these requirements, was installed by the General Electric Company. Fig. 1 is a map of the toll road showing the location of microwave stations and the manner in which they are interconnected.

Off-turnpike vhf base station sites were selected to provide the desired coverage of the road and its environs. These sites are on principal hills commanding a good view of the rolling terrain through which most of the Turnpike winds.

The microwave system links each base station with maintenance centers, police barracks, and administrative headquarters located in the heart of down town Boston.

Chicopee Falls is the central dispatch point and all vhf communication is relayed there via microwave. Here, dispatchers monitor all traffic and can transfer control of the system to any of the auxiliary control points at maintenance areas, police barracks, and administrative headquarters. During the day, two operators split the system; one acts as a dispatcher for the western half of the Turnpike; and one the eastern half.

The present vhf traffic load totals about 17,000 calls per month, or one call every $2^1/_2$ minutes. This will increase in the summer months with increased vehicular traffic.

In addition to the vhf system, separate microwave channels provide telephone service for maintenance, police, and administrative personnel. A high-speed party-line teletype loop was installed for the Turnpike police.

Vhf System

The vhf system layout is shown in Fig. 2. Each base station and control

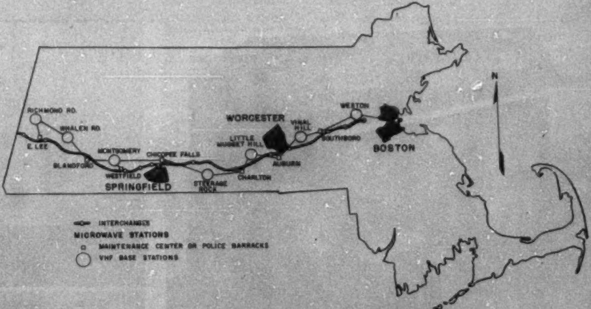


Fig. 1. Massachuetts Turnpike communication system

Paper 58-815, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 29, 1958.

RALPH F. LOWE is with the General Electric Company, Syracuse, N. Y.

The author wishes to acknowledge the work doneby A. C. Giesselman of the General Electric Company in designing the very-high-frequency system and control console.

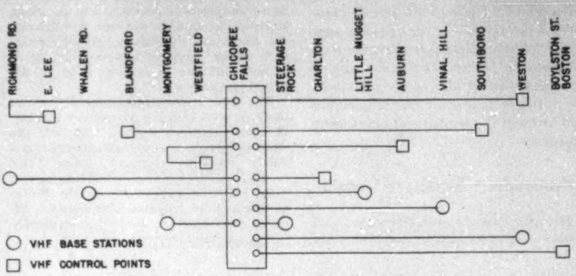

RICHMOND RD.
E. LEE
WHALEN RD.
BLANDFORD
MONTGOMERY
WESTFIELD
CHICOPEE FALLS
STEERAGE ROCK
CHARLTON
LITTLE MUGGET HILL
AUBURN
VINAL HILL
SOUTHBORO
WESTON
BOYLSTON ST. BOSTON
VHF BASE STATIONS
VHF CONTROL POINTS

Fig. 2. Vhf system microwave channel diagram

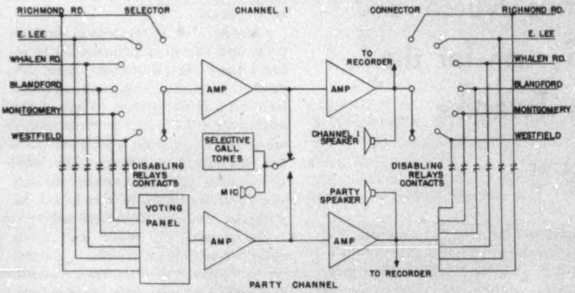

RICHMOND RD.
SELECTOR
CHANNEL I
CONNECTOR
RICHMOND RD.
E. LEE
WHALEN RD.
BLANDFORD
MONTGOMERY
WESTFIELD
TO RECORDER
AMP
AMP
SELECTIVE CALL TONES
CHANNEL I SPEAKER
DISABLING RELAYS CONTACTS
MIC
PARTY SPEAKER
VOTING PANEL
AMP
AMP
TO RECORDER
PARTY CHANNEL

Fig. 3. Simplified control console block diagram

point circuit is brought into the control console at Chicopee Falls on a separate microwave voice channel. Since a single 150-mc (megacycle) vhf frequency pair is shared by police, maintenance services, and administrators, flexibility is a prerequisite of the system. Selective calling of mobiles and interchanges, individual base station control, and a unique control console are combined to provide optimum performance.

Control Console

Fig. 3 is a simplified block diagram of a portion of the control console (only the western circuits are shown). The console contains three voice channels and two operating positions which can be seen in Fig. 4. Each of the vhf base station and control point circuits are connected in multiple to the party circuit and the operators station selector and connector banks shown in their home positions.

The system operates full duplex push-to-talk so that the control circuits, as well as the audio pairs are patched through the control console. Under quiescent conditions, circuits are connected to the party channel through the voting panel, as shown in Fig. 3. This unit serves as a base station selector. When an incoming call from a mobile is received on more than one base station receiver, only one signal is applied to the party channel.

The signal received on the party channel is retransmitted from all base stations and received at all control points. The dispatcher, after making an initial contact, may, at his discretion, transfer the call to channel 1 and free the party channel for other traffic. This is accomplished by the rotary selector and connector moving to the desired station position under the control of one of the push buttons in the vertical rows on either side of the operating position shown in Fig. 5. The disabling contacts open, removing the circuit from the party channel.

Fig. 6 shows the manner in which two vhf circuits may be interconnected through one of the control console channels. Operation of the station push button $S1$ starts relay $K1$ stepping through its interrupter contacts IC until the contact grounded through $S1$ is reached. The disabling relay associated with the position operates, disconnecting

Fig. 4. Massachusetts Turnpike vhf control console

Fig. 5. Control console operating position

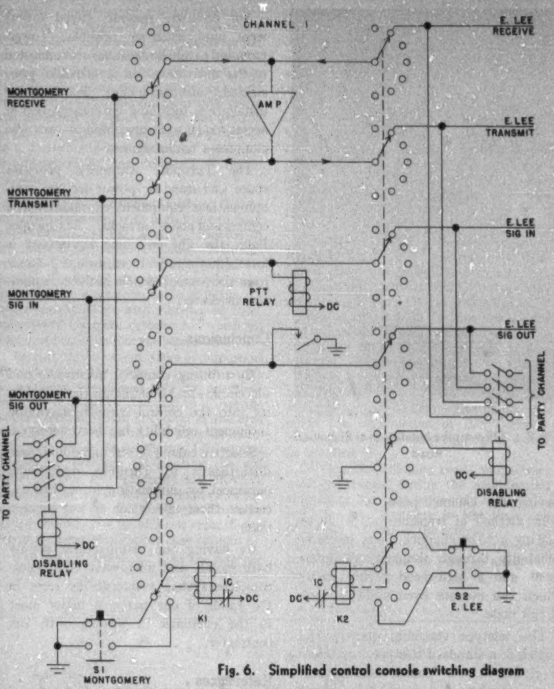

CHANNEL 1
E. LEE RECEIVE
MONTGOMERY RECEIVE
AMP
E. LEE TRANSMIT
MONTGOMERY TRANSMIT
E. LEE SIG IN
PTT RELAY
DC
MONTGOMERY SIG IN
E. LEE SIG OUT
MONTGOMERY SIG OUT
TO PARTY CHANNEL
TO PARTY CHANNEL
DISABLING RELAY
DC
DISABLING RELAY
DC
IC
DC
K1
IC
DC
K2
S2 E. LEE
S1 MONTGOMERY

Fig. 6. Simplified control console switching diagram

the station circuits from the party channel. Each level of *K*1 connects its associated circuits through the console channel. *S*2 operates the selector switch *K*2 in the same manner and both vhf circuits are connected together. At the completion of the call, the operator must return the selector and connector switches to their home positions by depressing the associated push buttons.

The disabling relays may also be controlled from switches located between the operating positions on the control console, as well as automatically by the selector and connector switches. This feature enables the operator to disconnect a base station from the system in the event that it interferes with normal operations.

A 3-channel tape recorder is provided to keep a record of all traffic carried on any of the console channels. The recorder is started each time a channel is keyed and can record three simultaneous conversations.

BASE STATIONS

Fifty-watt base stations are normally connected directly into a microwave channel but may be switched to operate as an independent mobile repeater. When an incoming signal energizes the carrier operated relay, a voting sequence is started which delays the keying of the microwave channel for a short time. The duration of the delay is inversely proportional to the strength of the incoming signal, the maximum time being about 400 milliseconds, and occurring on the weakest signals.

Generally, two base stations receiving a signal from the same mobile will key their associated microwave channels at slightly different times. The first signal received by the console is selected by the voting panel and any subsequent signals are muted.

Corner reflectors were selected as base station transmitting antennas because of their high front-to-back ratio and they were oriented to minimize phasing areas on the Turnpike.

SELECTIVE CALLING

Mobile and interchange receivers are muted unless the transmission is preceded by a short audio tone burst of the proper frequency.

The dispatchers at the control console can select any of the following groups:

1. Police
2. Maintenance
3. Interchanges
4. Individual administrators
5. All call

If the circuit is being used for police traffic, the mobile receiver in any other group will emit a busy tone when the operator removes the microphone from its hook switch.

INTERCHANGES

Each interchange is equipped with a fixed mobile-type station located in the utility building. There is an extension circuit so that calls can be taken in the central toll booth on a flush wall-mounted speaker and microphone combination.

Microwave System

The microwave system is comprised of 15 stations separated an average distance

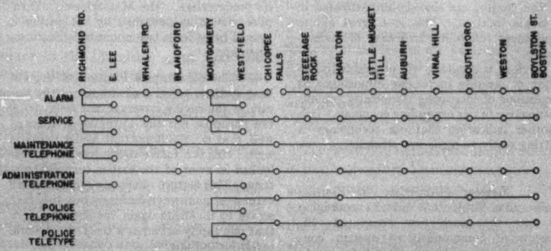

RICHMOND RD.
E. LEE
WHALEN RD.
BLANDFORD
MONTGOMERY
WESTFIELD
CHICOPEE FALLS
STEERAGE ROCK
CHARLTON
LITTLE MUGGET HILL
AUBURN
VINAL HILL
SOUTHBORO
WESTON
BOYLSTON ST. BOSTON
ALARM
SERVICE
MAINTENANCE TELEPHONE
ADMINISTRATION TELEPHONE
POLICE TELEPHONE
POLICE TELETYPE

Fig. 7. Microwave channel diagram exclusive of the vhf system

of 10 miles. The location of all stations was dictated by the communications requirements and vhf coverage.

A 2,000-mc time division microwave provides almost instantaneous selection and control of base stations by using the channel pulse as the signalling medium. Audio signals are transmitted by position modulation of the channel pulse.[1,2]

The antenna systems were designed to provide path margins which would yield an overall system propagation reliability of 99.6% for the worst month.[3]

Standby radio-frequency equipment is provided on all mainline circuits and an alarm channel provides an indication at Chicopee Falls of the condition of each station. A party line service channel telephone is located at each station to facilitate microwave and vhf maintenance.

Fig. 7 shows the microwave channels exclusive of the vhf channels.

Telephone Channels

The telephone channels are "party line" with dial selective ringing of all parties. Dial pulses are transmitted to all microwave stations by keying the channel pulse in accordance with the digits dialed. All selective ringing units follow the dialing but only one responds by ringing the extension dialed.

The selective ringing units installed on the administration telephone channel provide an additional feature, that of lockout and busy signal to all other parties when the circuit is in use.

Teletype Channel

The police teletype channel is a party channel linking teletype machines at all police barracks and the administrative headquarters in Boston. Teletype signals are transmitted over microwave by keying the channel pulse on and off. This method of transmission, while requiring a voice channel, offers increased reliability through simplicity of equipment and a high-speed keying circuit which can operate even under adverse circuit noise.

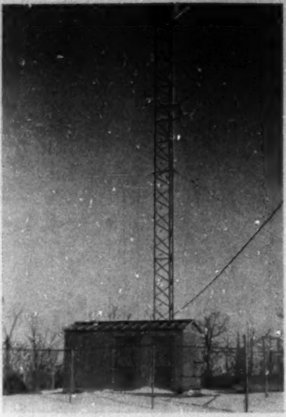


Fig. 8. Off-turnpike installation at Richmond Road

The teletype machines operate half duplex in a standard teletype loop which is closed at the Boston terminal. A message typed on any machine is received on the same machine via the loop.

Installations

Fig. 8 shows the off-turnpike installation at Richmond Road. Stand-by power is provided by a 5-kva air-cooled liquid petroleum-fueled engine generator installed in a separate room.

For economy reasons, guyed towers were used wherever possible. Tower-mounted parabolic antennas are connected to the microwave via semiflexible, pressurized transmission lines. The complete installation, including the construction of access roads, was supervised by the radio equipment manufacturer.

The Turnpike Authority provided space and stand-by power for the communications equipment at maintenance centers and police barracks. At Chicopee Falls, the vhf switching equipment is located in the radio room, a short distance from the control console in the communications center.

Conclusions

By utilizing standard microwave and vhf circuits and building all the complexity into the control console, maximum equipment reliability has been achieved.

Selective calling of vhf units minimizes disturbances to Turnpike operational personnel by eliminating from their receivers, those calls which do not concern them.

By having one manufacturer supply both mobile and microwave equipment, complete system responsibility rests in the hands of one party, a major asset to the customer in dealing with the contractor.

References

1. Quadriphase—A New Approach to Time-Division Multiplexing, W. E. Evans, R. F. Lowe. *AIEE Transactions (Electrical Engineering)*, vol. 74, pt. III, Oct. 1955, pp. 900–03.

2. The Quadriphase Microwave System, S. R. McConoughey. *Radio and Television News*, New York, N.Y., Electronic Engineering edition, Apr. 1955.

3. Microwave System Parameters for Reliable Communications, A. Alvira, I. T. Corbell. *AIEE Transactions*, vol. 74, pt. III, Dec. 1955, pp. 1100–07.

◆

Discussion

Paul C. Shedd (Newark College of Engineering, Newark, N. J.): This paper represents an excellent example of how special enterprises in modern communications can be specifically tailored and designed to meet situations which will not occur again in the same specific form.

A large number of different requirements and conditions are provided for with a minimum amount of fuss, and great flexibility is obtained, by combining standard circuits in a carefully developed plan.

Although the system extends practically across the state of Massachusetts between east and west, the most complex components of the control are sharply concentrated in a single location. This is a great aid to equipment reliability and ease of maintenance.

Work of this nature will undoubtedly become of steadily increasing importance as conditions of increasing complexity present themselves in our expanding economy. It is another indication that our technology is getting ready to meet new situations.

W. J. Kessler (University of Florida, Gainesville, Fla.): It is indeed encouraging to note that each new integrated vhf-microwave communications system constructed for turnpike use constitutes in one way or another a distinct improvement over its predecessors. The Massachusetts Turnpike system as described by the author is graced by at least two noteworthy features. They are:

1. The off-turnpike location of the vhf base stations to take advantage of the terrain in improving coverage not only along the Turnpike but the Turnpike environs as well, that is; the important feeder roads as well. This is a factor which has been overlooked in most of the earlier turnpike systems. This feature obviously requires more extensive microwave facilities but this fact serves to illustrate again the simple axiom that one rarely achieves a truly worthwhile result for nothing!

2. The use of a single vhf frequency pair

for police, maintenance, and administrative functions. It appears that this feature—which is certainly a meritorious one from the standpoint of vhf frequency conservation—is practical because of the use of selective tone signalling and the use of busy-channel indication in all mobile units when the police vhf channel is in use. Should this scheme prove fully practical over the long haul, it is a feature which is certainly worthy of consideration by the designers of future systems.

The author has referred briefly to the use of corner reflectors with high front-to-back ratios at the vhf base stations oriented to minimize phasing-area problems. Such vhf-overlap areas which are characterized by increased hash and noise in the mobile units, is one of the more formidable problems always encountered with integrated vhf-microwave turnpike systems for which no fool-proof solution has yet been found.

In fact this problem alone is important enough to be the topic of another paper. Concluding, one cannot help but speculate whether it might be practical to utilize the rolling terrain, which is typical of the area through which most of the Massachusetts Turnpike winds, to ameliorate further this phase-area problem by locating the vhf base stations and adjusting the effective radiated powers so that the high-elevation spots of the terrain along the Turnpike proper would serve to isolate or separate the coverage areas of adjacent base stations

The author points out that the teletype signals are transmitted by keying the channel pulse off and on. I can well appreciate the simplicity of this scheme but must admit that I do not grasp completely the reasons for the high noise-immunity properties suggested in the paper, since it appears that the full voice channel is not effectively utilized for the transmission of information-bearing signals. Unless of course the amplitude-modulation spectrum of the channel pulse under high-speed keying conditions extends throughout the entire voice channel. Perhaps this point will be clarified during the presentation.

Ralph F. Lowe: The discussion by Paul Shedd brings out an important point, that of the careful planning which must go into the engineering of a system of this type.

Because all the radio equipment was from one manufacturer, close co-operation was easily maintained between microwave and mobile application engineers.

The field project engineer supervised most of the application work before moving his residence to the field where he co-ordinated the construction of the Turnpike buildings in which radio equipment would be installed.

Off-turnpike buildings, roads, and towers were constructed during the manufacture and test of the radio equipment and the installation was complete in less than one year from the date of the contract.

Professor Kessler's speculation regarding the effect of terrain on phasing areas undoubtedly is the explanation for the short overlap areas observed on the Massachusetts Turnpike. Another factor, which was brought out during the presentation, is the affect of the relative phase of the audio signals modulating the interferring transmitters. By using individual microwave channels to connect the base stations the phase shift difference between them is minimized and no special correcting networks are required.

The teletype circuit actually uses a binary pulse code modulation system which should provide approximately 3 db signal-to-noise improvement over a pulse position modulation system.

This can be explained by the fact that the intrinsic noise shows up as incidental time modulation of the channel pulse due to the action of the clippers and reshapers following the receiver. The signaling circuit depends only upon the presence of a channel pulse in the gate and not its position in the gate or the fact that it may be changing position due to modulation or noise. In addition, the circuit need only detect a full-on or -off (binary) condition and full noise (no signal) does not have sufficient duty cycle to operate the relays.

Experience indicates that circuit noise causes more teletype "hits" when using frequency shift subcarriers than when using pulse code modulation.

Even though the amplitude-modulation spectrum of the teletype signal does not extend very far over the voice channel, the space is unavailable for other circuits.

Tropospheric-Scatter System Design

LEANG P. YEH
NONMEMBER AIEE

THIS paper attempts to present a comprehensive procedure, from an approach partly experimental and partly theoretical for the design of tropospheric-scatter communication systems, using single side band (SSB) or frequency modulation (FM) and frequency division multiplex (FDM).

Since the art is still young, controversial views and discrepancies are found elsewhere in the literature. But from the system design point of view, some means must be provided to determine the parameters and to predict fairly well the performance from such parameters. This paper is written, therefore, with this goal in mind.

The methods as given are by no means the absolute choice. It is hoped that by the same approach, a unique method could be obtained and agreed upon someday by all concerned.

System Design

System design for an FDM operation is based on the required signal-to-total-noise power ratio per channel at the receiver output.

The total noise in any FDM radio system has two origins:

1. Intermodulation noise due to nonlinearity of the system components.
2. Other noises such as thermal noise inherent in the receiver, and external noises such as cosmic noise and man-made noise received by the antenna.

The treatment of intermodulation noise in system design is discussed in a separate paper under preparation. Only the second type of noise is covered here. In the region of 300 to 4,000 megacycles or above, extensively used in tropospheric scattering, the thermal noise of the receiver predominates at isolated locations where the terminals are likely to be located. This second type of noise will, therefore, consist of thermal noise only.

Table I outlines the general procedure for the design of a tropospheric-scatter system, based on the minimum usable signal-to-thermal-noise power ratio per channel at the receiver output.

MINIMUM-CHANNEL SIGNAL-TO-THERMAL NOISE POWER RATIO FOR USABLE COMMUNICATION

It seems that there is as yet no universal agreement on the minimum usable channel signal-to-noise power ratio. Figures, available for circuits carrying speech, from various sources are given in Table II.

Paper 58-816, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 12, 1958; made available for printing April 17, 1958.

LEANG P. YEH is with Westinghouse Electric Corporation, Baltimore, Md.

Permission to release the information in this paper by Westinghouse Electric Corporation is greatly appreciated. The author is also indebted to N. B. Tharp, E. L. Smith, and J. B. Murray for their encouragement, guidance, and helpful suggestions.

TABLE I - SYSTEM DESIGN

1. Minimum Signal-to-Thermal noise power ratio per channel - Min$[\frac{S_0}{N_0}]$

$$\text{Min}\ [\frac{S_0}{N_0}] = 25 \text{ db flat (0 dbm test tone)}$$
$$= 15 \text{ db flat (average speech)}$$

2. Minimum Carrier-to-Thermal noise power ratio - Min$[\frac{P_R}{N_A}]$

	SSB	FM
$[\frac{S_o}{N_o}]$	$[\frac{P_R}{N_a}] + G_1 - L$	$[\frac{P_R}{N_a}] + G_1 + G_{FM} - L$
$[\frac{P_R}{N_a}]$	Peak Carrier-to-average thermal noise power ratio in db	average carrier-to-average thermal noise power ratio in db
G_1	Filter imp. = $10 \log_{10} \frac{B_{if}}{B_c}$ in db	Filter imp. = $10 \log_{10} \frac{B_{if}}{2B_c}$ in db
B_{if}	Receiver if bandwidth = f in cps	Receiver if bandwidth = $2(F_d+2f)$ in cps
B_c	Channel bandwidth = 3000 cps	Channel bandwidth = 3000 cps
f	Top multiplex frequency in cps	Top multiplex frequency in cps
F_d	-	Peak freq. dev. for system in cps
G_{FM}	-	Noise imp. = $20 \log_{10} \frac{F_d}{f}$ in db
L	Channel loading factor = peak multiplex level to rms channel level = peak-to-rms multiplex power ratio (10 db-0.1% of time) + rms multiplex-to-rms channel power ratio (adjusted Holbrook and Dixon)	
Threshold		Threshold $[\frac{P_R}{N_a}]$ = 10 db Threshold $[\frac{S_o}{N_o}] = 10+G_1+G_{FM} - L$

3. Median Carrier-to-Thermal noise power ratio - Med$[\frac{P_R}{N_A}]$

	SSB	FM
$[\frac{P_R}{N_A}]$	$10 \log_{10} P_R - 10 \log_{10} N_A \doteq 10 \log_{10} P_T - L_T - 10 \log_{10} N_A$	
P_R	Peak recvd. carrier power in watts	Ave. recvd. carrier power in watts
P_T	Peak trans. carrier power in watts	Ave. trans. carrier power in watts
L_T	Total System loss in db = $L_S \pm L_{CF} + L_{FS} - G_T - G_R + L_C + L_A$	
L_S	Experimental Scatter loss = 57 + .15 (d*mi 100) in db	
L_{CF}	Median Scatter Loss correction factor in db due to terrain(Kurihara)[1]	
L_{FS}	Free Space loss in db = $36.6 + 20 \log_{10} d_{mi} + 20 \log_{10} f_{mc}$*	
G_T	Plane wave trans. ant. gain in db $\doteq 8 + 20 \log_{10} \frac{D^*}{\lambda}$	
G_R	Plane Wave rec. ant. gain in db $\doteq 8 + 20 \log_{10} \frac{D^*}{\lambda}$	
L_C	Median aperture medium coupling loss in db = (Booker and DeBettencourt)[2]	
L_A	Total Feeder and junction losses = $5 \log_{10} f_{mc} - 10$	
N_A	Thermal Noise power = $FKTB_{if}$ in watts	
F	Receiver Noise figure in ratios	
K	Boltzmann's constant = 1.38×10^{-23} joules/°K	
T	Temperature = 290°K	
B_{if}	IF bandwidth = f	IF bandwidth = $2(F_d + 2f)$

*dmi - path length in statute miles f_{mc} - carrier frequency in mcps

D - antenna aperture in ft λ - wavelength in ft

4. Fading Margin and System Reliability

$$\text{Fading margin in db} = \text{Med.}[\frac{P_R}{N_A}] - \min[\frac{P_R}{N_A}] + G_c$$

G_c = combiner over selector diversity gain

The System reliability can be read off from long-term instantaneous signal distributions (probability combination of short-term and long-term distributions).

Table II. Minimum Usable Signal-to-Thermal-Noise Power Ratio

Source	Average-Speech-to-Average-Noise Ratio, Db	Peak-Speech-to-Average-Noise Ratio, Db	Remarks
Barsis[3]	11		88% reliability
Morrow[4]	15	25*	90% articulation without compandor
Morrow[4]	10	20*	90% articulation with compandor
C.C.I.R.[5]	15	21†	90% intelligibility without compandor

* Peak speech is assumed to be 10 db above average speech.

† Peak speech is assumed to be 6 db above average speech.

It is understood that these figures are also good for other kinds of signals, except perhaps facsimile.

It is suggested that a minimum of 15-db (decibel) channel average-speech-to-average-thermal-noise power ratio (without compandor) is considered usable.

Since the channel signal-to-thermal-noise power ratio is usually specified for 0-dbm (decibels referred to one milliwatt) test tone, the minimum usable ratio becomes 25 db on this basis, if the average speech level is assumed to be about −10 dbm.[6]

MINIMUM CARRIER-TO-THERMAL-NOISE POWER RATIO FOR USABLE COMMUNICATION

Once the minimum-channel signal-to-noise power ratio for usable communication is established, the corresponding minimum carrier-to-thermal-noise power ratio could be obtained if the number of multiplex channels and the modulation techniques are known.

For FM systems, there is also the threshold effect. The average-carrier-to-average-thermal-noise power ratio at practical threshold has been established as 10 db.

Therefore, both the minimum (S_0/N_0) and threshold (P_R/N_A) should be considered in designing FM systems. From

Vol.
Pt. 1
1958

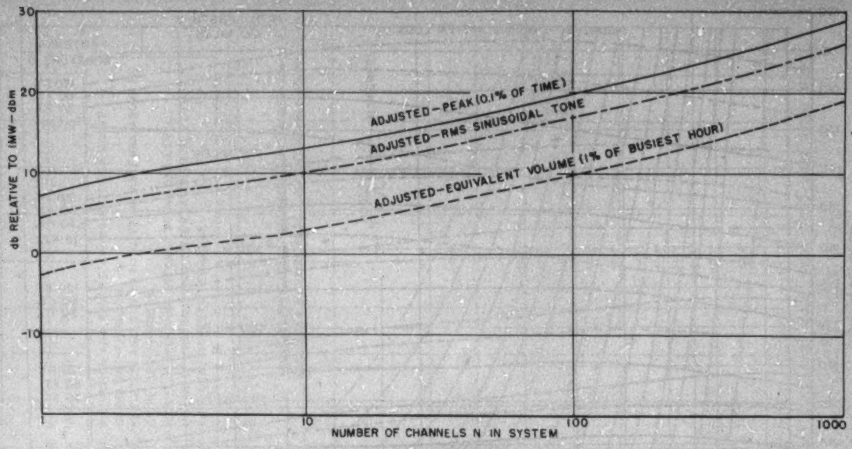


Fig. 1. Channel loading factor

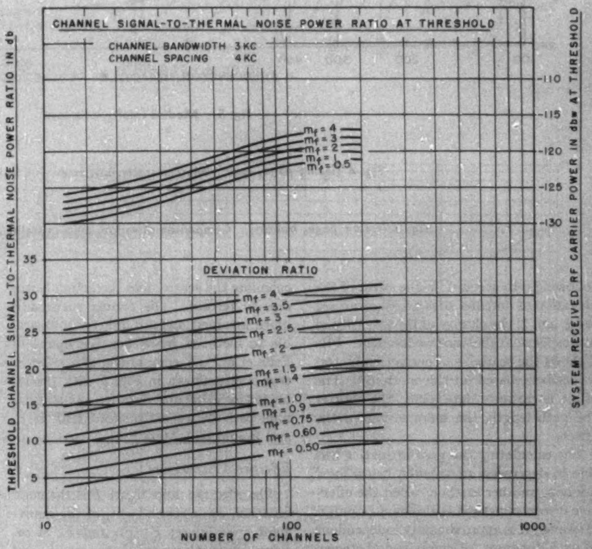


Fig. 2. Threshold channel signal-to-thermal-noise power ratio

them the deviation ratio is usually determined.

Fig. 1 gives the channel loading factor (adjusted Holbrook and Dixon[7]) and Fig. 2 gives the channel signal-to-thermal-noise power ratio versus number of channels at various deviation ratios.

MEDIAN CARRIER-TO-THERMAL-NOISE POWER RATIO

The median received carrier power is the difference between the transmitter power and the median total system loss. The thermal-noise power may be calculated if both the noise figure and i-f (intermediate-frequency) bandwidth of the receiver are known.

Median Total System Loss

Fig. 3 gives the median total system loss for smooth-earth geometry. It is plotted with total loss versus frequency, path distance, and antenna aperture. It may be used to determine the optimum frequency, which is the one that gives minimum loss.

Terrain Effect

Since the majority of experimental data used for deriving the median total system loss was obtained from scatter geometry, fairly similar to smooth earth, the effect of terrain must be considered in dealing with nonspherical-earth scatter geometry.

The effect of terrain, defined as scatter-loss correction factor, could be a loss or gain, depending on whether the antenna is elevated or depressed in elevation relative to the local horizon. (This is different from the so-called radio horizon; see Fig. 5.) It is also statistical in nature, but the median value is of importance. Several methods have been advocated to calculate the median value. Kurihara's method[1] is adopted because it takes the

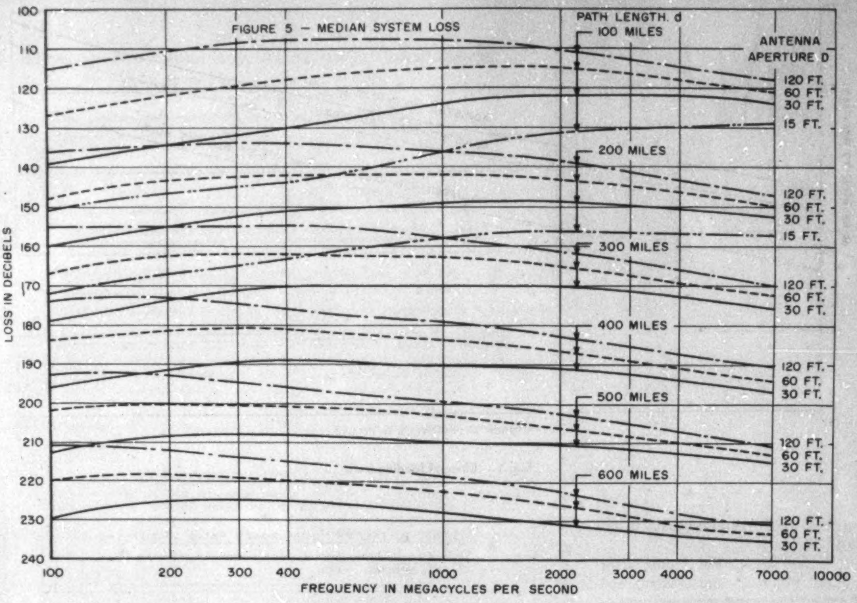


Fig. 3. Median total system loss

Fig. 4 (facing page, top). Median scatter-loss correction factor due to terrain

Fig. 5 (facing page, bottom). Comparison of various median scatter-loss correction factors

following into account: 1. the effect of the elevation of antenna sites above mean sea level, and 2. the change of the intensity of atmospheric turbulence with height. Fig. 4 shows the scatter-loss correction factor[8] in decibels, based on this method.[8] The effect is distance dependent, the shorter the path length, the more serious is the effect.

For calculating the gain in scatter loss due to depression of antenna below local horizon, another method called the effective distance method is also very popular. However, it is approximately independent of path distance. A comparison between this method and Kurihara's is shown in Fig. 5.

The effective distance method cannot be applied to cases when the antennas are elevated above local horizons to clear obstacles along the path. However, there are two other methods advocated by Norton, et al.,[9] and Friis, et al.[10] to calculate the scatter-loss correction factor due to terrain. The former method is more involved and not directly comparable to methods by either Kurihara[1] or Friis et al.[10] A comparison of the latter two is also shown in Fig. 5 with the reported experimental data for 188-statute-mile path between Crawford Hill, N.J., and Round Hill, Mass.[11]

Effective Noise Figure

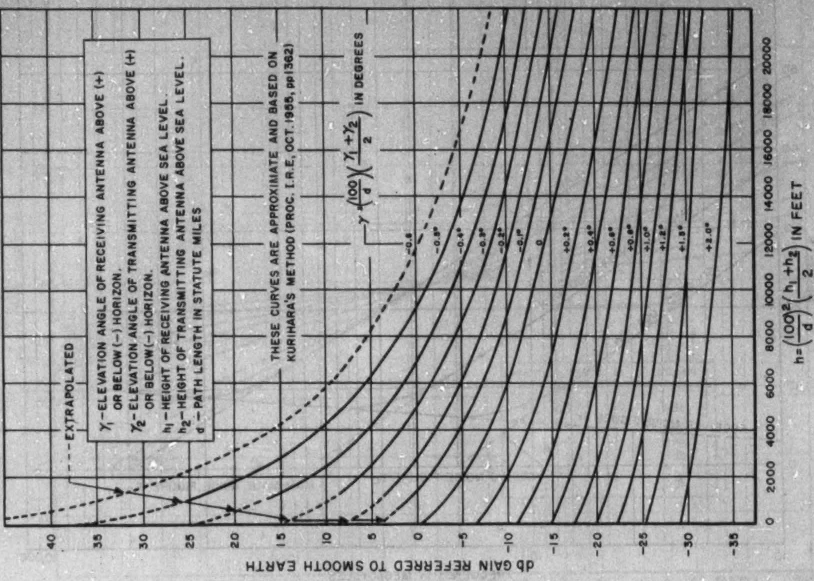
The effective noise figure F is the combination, on a power basis, of the equivalent noise figures F_r, F_c, and F_n of receiver thermal noise, cosmic noise, and man-made noise respectively. In the event that the actual value is not available, approximate values may be obtained from Fig. 6.

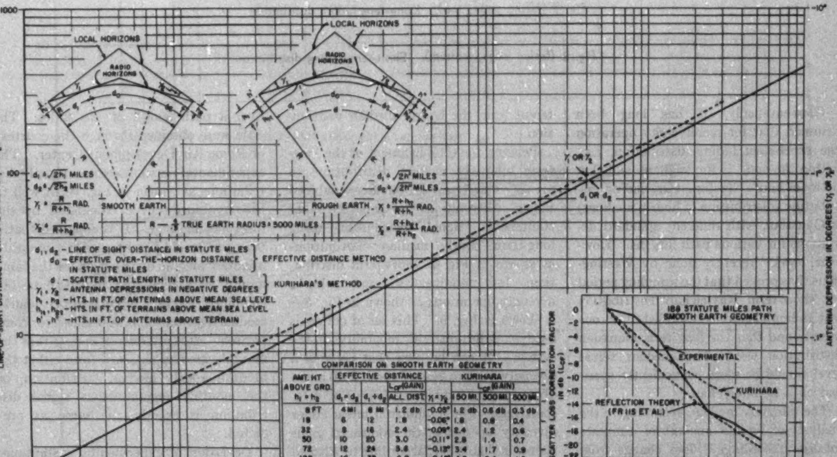
FADING DISTRIBUTION

Heretofore, the fading is arbitrarily subdivided into two components because of their natures, fast and slow, in order to facilitate the mathematical analysis. The slow fading is believed to be caused by the change in tropospheric conditions, such as the change of index of retraction of the atmosphere and the fast fading by multipath transmissions.

Practically, the two are inseparable because the signal received is always the combination of both, which may be called the instantaneous signal. Therefore, if the instantaneous signal distributions of the hour, the day, the month, or the year are known, the system reliability for a certain fading allowance could be predicted fairly precisely for whatever period is intended.

The instantaneous signal distributions could be obtained experimentally. If an experimental curve is not available, the probability combination of short-term and long-term fading distributions may be used instead.





COMPARISON ON SMOOTH EARTH GEOMETRY

AMT. HT ABOVE GRD. $h_1 = h_2$	EFFECTIVE DISTANCE			KURIHARA			
	$d_1 = d_2$	$d_1 + d_2$	L_{CF} (GAIN) ALL DIST.	$\gamma_1 = \gamma_2$	L_{CF} (GAIN) 150 MI.	300 MI.	600 MI.
8 FT	4 MI	8 MI	1.2 db	−0.05°	1.2 db	0.6 db	0.3 db
18	6	12	1.8	−0.06°	1.8	0.9	0.4
32	8	16	2.4	−0.09°	2.4	1.2	0.6
50	10	20	3.0	−0.11°	2.8	1.4	0.7
72	12	24	3.6	−0.13°	3.4	1.7	0.9
128	16	32	4.8	−0.18°	4.8	2.4	1.2
200	20	40	6.0	−0.22°	6.0	3.0	1.8
450	30	60	9.0	−0.34°	9.2	4.6	2.3

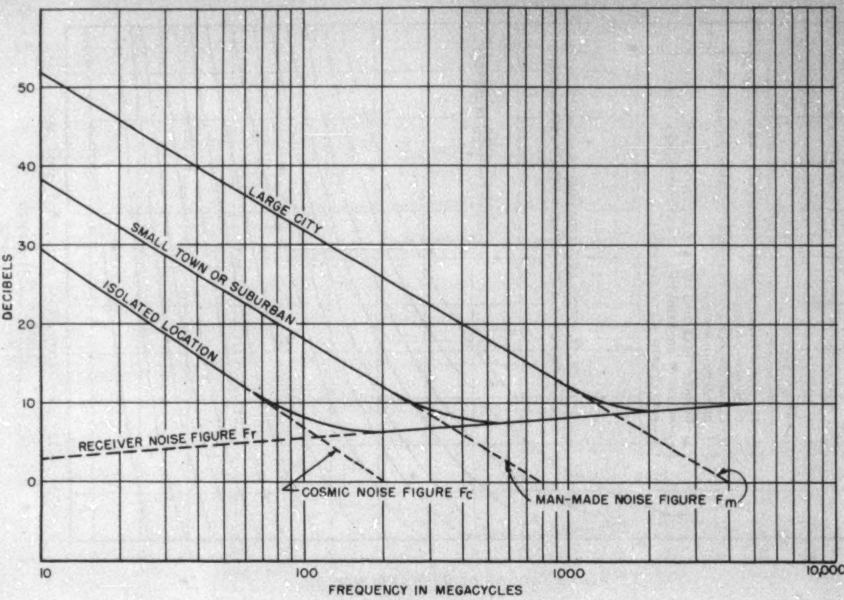


Fig. 6. Effective noise figure

Fig. 7 (facing page, top). Long-term fading distribution

Fig. 8 (facing page, bottom). Short-term fading distribution

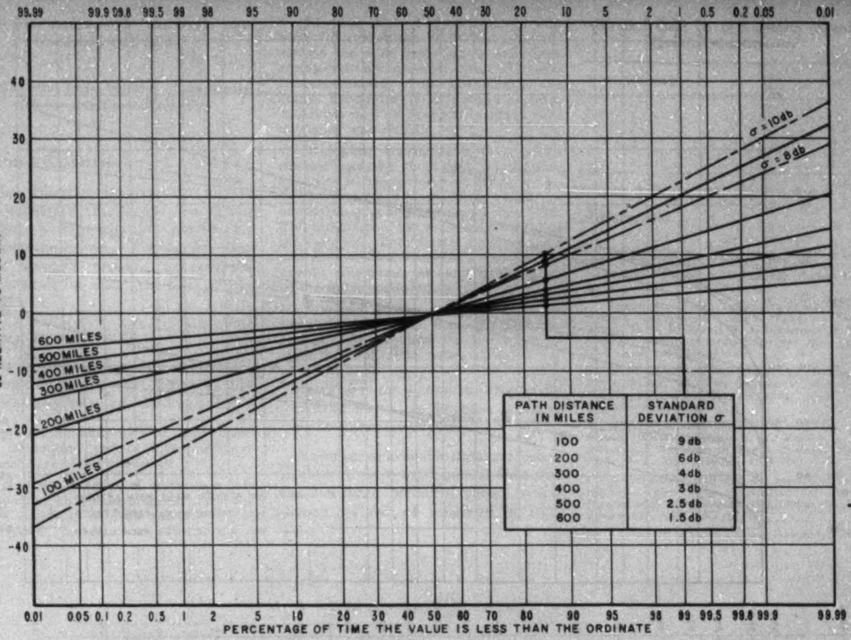
Conventionally, it has long been assumed that for nondiversity operation the short-term fading distribution, e.g., within the hour, is Rayleighan; and that between hours, the long-term fading distribution (distribution of hourly medians) is Gaussian, with different standard deviations for various path lengths. However, work[12] recently done by Westinghouse has proved that the sampling period for short-term Rayleighan Distribution fast fading should be established as one minute and that the long-term Gaussian distribution slow-fading median signal level should be more conveniently specified on a minutely basis.

The minutely median, according to the analysis of limited amount of 1-minute totalizer recordings,[12] does change from minute to minute. The distribution also approaches Gaussian. It is believed that if enough data are taken on a long-term basis (monthly or yearly), minutely median distribution should not be much different from the hourly median distribution.

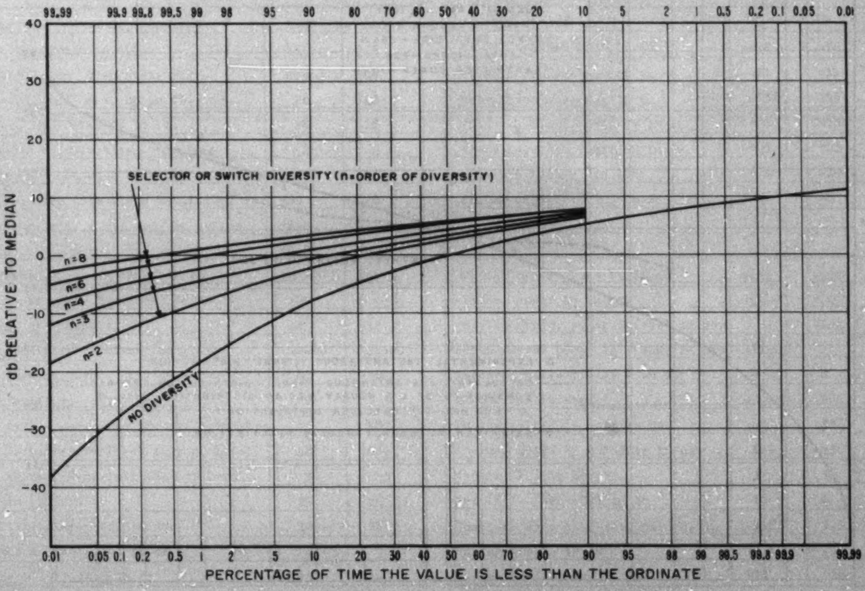
Pending the establishment of the long-term minutely median distributions, a set of long-term probability-combined fading distributions or long-term instantaneous signal distribution—using long-term hourly median distributions as shown in Fig. 7, short-term distributions for nondiversity, and selector diversity operations as shown in Fig. 8—is shown in Fig. 9. This set of curves is used to determine the reliability of the system. It can be seen that there is a diversity median gain of about 2 db and 4 db for order of diversity $(n)=2$ or 4 respectively. For combiner or linear adder diversity operation, the additional gains shown in Table III, according to Altman and Sichak,[13] should be added.

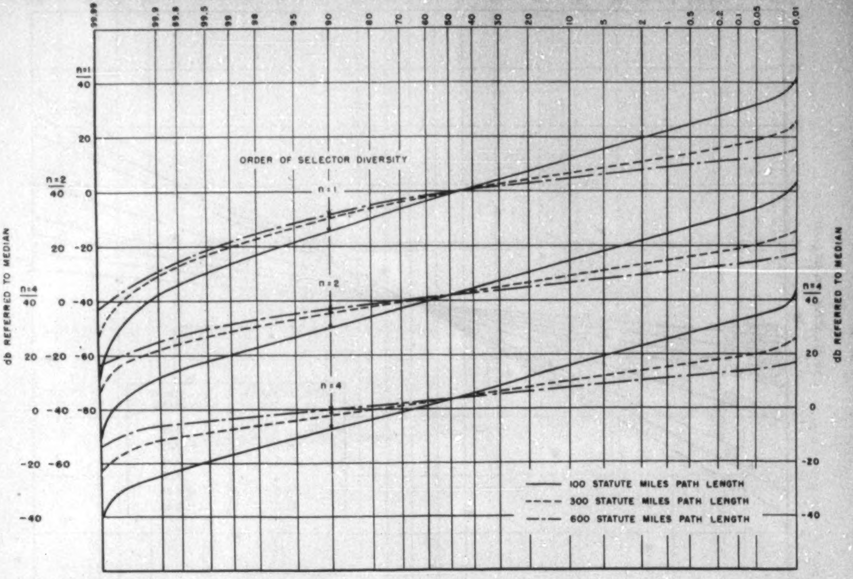
Westinghouse has analyzed 293 hours of the data between January and February 1957 on the Baltimore, Md., and Verona, N.Y., link operating at 960 megacycles over a path length of 268 miles. The data were obtained through the courtesy of Rome Air Development Center. The experimental instantaneous signal distributions as shown in Fig. 10 fit well to the combined curve calculated by the foregoing method. It is expected that a better fit would be obtained if a monthly minutely median distribution for the same month were available to calculate the combined curve. Experimental hourly median distribution is also shown for comparison. It is interesting to note that the monthly median obtained from either the monthly-hourly median distribution, or the monthly instantaneous signal distribution is exactly the same as predicted.

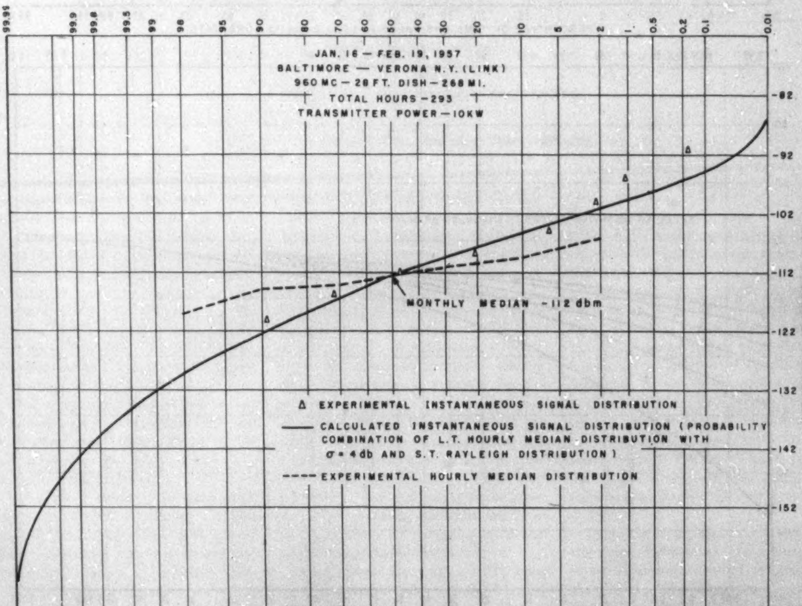
Deviation of experimental instantaneous signal distribution from calculations at the strong signal region indicates that there is signal enhancement due to external effects such as aircraft, trapping, or ducting, etc.



PATH DISTANCE IN MILES	STANDARD DEVIATION σ
100	9db
200	6db
300	4db
400	3db
500	2.5db
600	1.5db




99.99
99.9
99.8
99.5
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.05
0.01
ORDER OF SELECTOR DIVERSITY
n=1
n=2
n=4
n=1 40
20
n=2 40
0
20 -20
n=4 40
0 -40
20 -20 -60
0 -40 -80
-20 -60
-40
db REFERRED TO MEDIAN
n=4 40
20
0
-20
-40
db REFERRED TO MEDIAN
100 STATUTE MILES PATH LENGTH
300 STATUTE MILES PATH LENGTH
600 STATUTE MILES PATH LENGTH


99.99
99.9
99.8
99.5
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.01
JAN. 16 — FEB. 19, 1957
BALTIMORE — VERONA N.Y. (LINK)
960 MC — 28 FT. DISH — 268 MI.
TOTAL HOURS — 293
TRANSMITTER POWER — 10 KW
MONTHLY MEDIAN -112 dbm
-82
-92
-102
-112
-122
-132
-142
-152
RECEIVED POWER IN dbm
Δ EXPERIMENTAL INSTANTANEOUS SIGNAL DISTRIBUTION
CALCULATED INSTANTANEOUS SIGNAL DISTRIBUTION (PROBABILITY COMBINATION OF L.T. HOURLY MEDIAN DISTRIBUTION WITH σ = 4 db AND S.T. RAYLEIGH DISTRIBUTION)
EXPERIMENTAL HOURLY MEDIAN DISTRIBUTION

Vol.
Pt. 1
1958

Fig. 9 (facing page, top). Long-term combined fading distribution

Fig. 10 (facing page, bottom). Experimental instantaneous signal distribution

SYSTEM RELIABILITY

The difference between the median and minimum (P_R/N_A) is the fading margin. System reliability can be read off from the curves in Fig. 9. For combiner diversity system, the combiner over selector diversity gain G_c should be added.

SAMPLE DESIGN

A sample design is shown in Table IV. A comparison between SSB and FM systems may be summarized as follows:

1. System reliability—For the same average transmitted power single side-band systems have higher reliability.

2. Threshold effect—FM system can only be operated above the practical threshold, below threshold it fails completely. In SSB system, there is no such effect.

3. I-f bandwidth—FM systems require wider i-f bandwidth. In order to accommodate the wide band, it is sometimes necessary to use larger antennas. In some instances because of the limitation of the usable bandwidth in tropospheric-scatter transmission, wide bandwidth simply cannot be accommodated.

4. Multipath transmissions—For multichannel operation, multipath transmission will cause frequency selective fading in SSB systems which can be overcome by the use of multiple-pilot amplitude regulators. With FM, the effects of frequency selective fading are complex. Both intermodulation noise and channel level instabilities can be produced. The former puts another limitation on the usable bandwidth.

5. Spectrum economy—As shown in the sample design, for the same base-band frequency, SSB required far less spectrum space than FM (48 kc versus 600 kc).

COMPARISON OF PREDICTED RECEIVED SIGNAL LEVEL BY PRESENT METHOD WITH EXPERIMENTAL DATA

A number of available data on received signal level from experimental links with the calculated predicted values by the present method are tabulated in Table V.

Table III. Gain G_c of Combiner Diversity Over Selector Diversity

Order of Diversity	Gain, Db
1	0
2	1.5
4	3.0
6	4.0
8	5.0

References

1. TRANS-HORIZON RADIOWAVE PROPAGATION OVER A NON-SPHERICAL EARTH, Y. Kurihara. *Report EE-229*, Electrical Engineering, Cornell University, Ithaca, N. Y., 1954.

2. THEORY OF RADIO TRANSMISSION BY TROPOSPHERIC SCATTERING USING VERY NARROW BEAMS, H. G. Booker, J. T. deBettencourt. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Mar. 1955, pp. 281–90.

3. DATA FROM CANADIAN SCATTER PROJECTS, A. P. Barsis. National Bureau of Standards, Boulder Laboratory, Boulder, Colo., Sept. 1956 (private communication).

4. SINGLE SIDEBAND TECHNIQUES IN UHF LONG-RANGE COMMUNICATIONS, W. E. Morrow, et al. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Dec. 1956, pp. 1854–73.

5. *Documents*, VIIIth Plenary Assembly, International Consultative Committee on Radio, International Telecommunication Union, Geneva, Switzerland, vol. I, 1957, p. 160.

6. *Documents*, XVIIth Plenary Assembly, Inter-

TABLE IV - SAMPLE CALCULATIONS

PROCEDURE	SSB	FM
Carrier frequency f in mcps	2000	2000
Wavelength λ in ft	0.5	0.5
Path length d in stat. mile	200	200
Trans. ant. dia. D_T in ft.	28	28
Rec. ant. dia. D_R in ft.	28	28
Order of diversity - **combiner**	2	2
Trans. Ant. Site above M.S.L.(h_2) in ft.	450	450
Rec. Ant. Site above M.S.L.(h_1) in ft.	500	500
Trans. ant. elev. (r_2) in degree	+1.0	+1.0
Rec. ant. elev. (r_1) in degree	+0.5	+0.5
Trans. power P_T (Ave) in KW	10	10
Trans. power P_T (Peak) in KW	100	
No. of channels (n')	12	12
Multiplex frequency in kcps	12-60	12-60
Top mod. freq. (f_t) in kcps	60	60
Deviation Ratio (m_f)		3
System peak freq. dev.(F_d) in kcps		180
IF bandwidth (B_{if}) in kcps	48	600
Channel bandwidth (B_c) in kcps	3	3
Min $[\frac{S_o}{N_o}]$ in db	25	25
Filter Imp. (G_i) in db	12	20
Noise Imp. (G_{FM}) in db		9
Loading factor in db	-14	-14
Min. $[\frac{P_R}{N_A}]$ in db	27	10
Median Scatter Loss L_S in db	-72	-72
Terrain correction L_{CF} in db	-11	-11
Free Space loss L_{FS} in db	-151	-151
Plane Wave Ant. Gain G_T in db	+43	+43
Plane Wave Ant. Gain G_R in db	+43	+43
Aperture Loss L_c in db	-10	-10
Feeder Loss, etc. L_a in db	-4	-4
Median System Loss L_T in db	-162	-162
Trans. power P_T in dbw	+50	+40
Median Recvd. power P_R(Ave) in dbw		-122
Median Recvd. power P_R (Peak) in dbw	-112	
Receiver noise figure F in db	8	8
KT per cps in dbw	-204	-204
10 $\log_{10}$ (B_{if} in cps) in db	47	58
Thermal noise N_A = $FKTB_{if}$ in dbw	-149	-138
Median $[\frac{P_R}{N_A}]$ in db	37	16
Med $[\frac{P_R}{N_A}]$ - Min $[\frac{P_R}{N_A}]$ in db	10	6
Combiner over Sel. div. Gain G_c in db	1.5	1.5
Fading Margin in db	11.5	7.5
System Reliability %	97%	91%

TABLE V

A. PREDICTED YEARLY MEDIAN RECEIVED SIGNAL LEVEL

SOURCE	BELL TEL. LAB., INC.		MIT LINCOLN LABORATORY[11]			NBS[14]	
LINK	Pharsalia, N.Y. - Crawford Hill, N. J.		Round Hill Cole Lab, N.J.		Crawford Hill Round Hill, Mass.	Cheyenne Mt. to Fayetteville	Cheyenne Mt. to Anthony
Carrier frequency f	4110 mc	4110 mc	385.5 mc	3670 mc	505 mc	100 mc	192.8 mc
Wave length λ	0.244 ft.	0.244 ft.	2.530 ft.	0.273 ft.	0.198 ft.	10 ft.	5.2 ft.
Path length d	171 mi.	171 mi.	184 mi.	188 mi.	188 mi.	617.7 mi.	393.5 mi.
Trans. ant. dia D_T	10 ft.	10 ft.	28 ft.	28 ft.	28 ft.	--	--
Rec. ant. dia. D_R	60 ft.	8 ft.	28 ft.	28 ft.	28 ft.	10 ft.	--
Trans. Power P_T	30 w.	30 w.	5 kw	400 kw	400 w.	2 kw.	2 kw.
P_T in dbm	+45 dbm	+45 dbm	+67 dbm	+86 dbm	+56 dbm	+63 dbm	+63 dbm
Plane wave ant. Gain $G_T + G_R$	+93 db	+76 db	+58 db	+95 db	+99 db	+30 db	+22 db
Med. Scatter Loss L_s	-68 db	-68 db	-69 db	-70 db	-70 db	-134 db	-102 db
Free Space Loss L_{FS}	-155 db	-155 db	-132 db	-153 db	-156 db	-133 db	-135 db
Med. Aperture loss L_c	-14 db	-2 db	0 db	-11 db	-12 db	0 db	- db
Feeder Loss L_a	-1 db	-1 db	-3 db	-1 db	-2 db	-2 db	-2db
Total System loss L_T	-145 db	-150 db	-146 db	-140 db	-141 db	-239 db	-217 db
Recd. Power P_R	-100 dbm	-105 dbm	-79 dbm	-54 dbm	-85 dbm	-176 dbm	-154 dbm
Rec. ant. elev. r_1	0	0	-0.23°	0	0	-0.09°	-0.09°
Trans. ant. elev. r_2	0	0	-0.16°	0	0	-0.74°	-0.74°
Rec. ant. ht. above M.S.L. h_1	370 ft.	370 ft.	221 ft.	53 ft.	53 ft.	1363 ft.	1344 ft.
Trans. ant. ht. above M.S.L. h_2	1880 ft.	1880 ft.	68 ft.	389 ft.	389 ft.	8805 ft.	8855 ft.
Rec. ant. ht. above terrain h_{1t}	--	--	--	--	--	38 ft.	39 ft.
Trans. ant. ht. above terrain h_{2t}	--	--	--	--	--	2270 ft.	2320 ft.
Scatter loss cor. L_{CF}	-1 db	-1db	+3 db	0 db	0 db	+3 db	+6 db
Net recd. power P_R	-101 dbm	-106 dbm	-76 dbm	-54 dbm	-85 dbm	-173 dbm	-148 dbm

B. EXPERIMENTAL MEDIAN SIGNAL LEVEL

Worst Month	-119 dbm (3/56)	-121 dbm (3/56)	-86 dbm (1/55)	-65 dbm (12/54	-95 dbm (1/54)	--	--
Yearly	-103 dbm (9/55 - 9/56)	-108 dbm (9/55 - 9/56)	-75 dbm (1/55-12/55)	-55 dbm (7/53-3/55)	--	--	--
Best Month	-95 dbm (9/56)	-100 dbm (9/56)	-65 dbm (7/55)	-39 dbm (8/53)	-82 dbm (8/54)	-168 dbm* (8/52)	-141 dbm* (8/52)

* These figures were obtained indirectly from the measured median basic transmission loss (L_{bm}) obtained by NBS people, see reference 14.

national Consultative Committee on Telephony, Geneva, Switzerland, vol. I, 1954, p. 198.

7. LOAD RATING THEORY FOR MULTICHANNEL AMPLIFIERS, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., vol. 18, 1939.

8. TROPOSPHERIC SCATTER SYSTEMS, L. P. Yeh. *Technical Report no. 3*, UHF Section, Communication Equipment Engineering, Westinghouse Electronics Division, Baltimore, Md., Apr. 1957.

9. THE USE OF ANGULAR DISTANCE IN ESTIMATING TRANSMISSION LOSS AND FADING RANGE FOR PROPAGATION THROUGH A TURBULENT ATMOSPHERE OVER IRREGULAR TERRAIN, K. A. Norton, et al. *Proceedings*, Institute of Radio Engineers, Oct. 1955, pp. 1488–1526.

10. A REFLECTION THEORY FOR PROPAGATION BEYOND THE HORIZON, H. T. Friis, et al. *Bell System Technical Journal*, New York, N.Y., May 1957, pp. 627–44.

11. INVESTIGATION OF ANGULAR SCATTERING AND MULTIPATH PROPERTIES OF TROPOSPHERIC PROPAGATION OF SHORT RADIO WAVES BEYOND THE HORIZON, J. H. Chisholm, et al. *Technical Report no. 84*, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., June 1955, p. 107, Fig. 72.

12. BASIC ANALYSIS ON THE CONTROLLED CARRIER OPERATION OF TROPOSPHERIC SCATTER COMMUNICATION SYSTEMS, L. P. Yeh. *Convention Record*, Institute of Radio Engineers, pt. I, 1958.

13. A SIMPLIFIED DIVERSITY COMMUNICATION SYSTEM FOR BEYOND-THE-HORIZON LINKS, F. J. Altman, W. Sichak. *Transactions*, Professional Group on Communication Systems, Institute of Radio Engineers, vol. CS-4, no. 1, Mar. 1956, pp. 50–55.

14. LONG DISTANCE TROPOSPHERIC PROPAGATION DATA OVER CHEYENNE MOUNTAIN PATHS, A. P. Barsis. *Report no. 5072*, National Bureau of Standards, Washington, D. C., May 7, 1957.

Radar Interference to Microwave Communication Services

R. D. CAMPBELL
MEMBER AIEE

IT IS THE PURPOSE of this paper to discuss the growing problem of interference between radar transmitters and microwave communication systems and the urgent need for the application of measures to enable these two radio services to coexist in the microwave spectrum. The problem basically is a difficult one since the radar transmitters use power outputs measured in megawatts, whereas the microwave transmitters normally use powers of 10 watts or less. The Bell System became aware of this problem as the result of a number of cases of interference to its microwave communication systems from radar operations. Two general types of interference are involved: 1. that resulting from the operation of radars directly on fundamental frequencies used for communication services, and 2. that due to the generation and radiation by the radars of spurious frequencies including harmonics which fall in bands and on frequencies used for communication services. The former type of interference can be eliminated through control of the frequency assignments for both radar and communication purposes. The latter type, however, is not so readily disposed of.

While a number of cases of radar interference have been experienced over the years since World War II, it was not until late 1954 that the seriousness of the problem was fully recognized in the telephone industry. During that year, severe radar interference was experienced to a *TD-2* microwave system operated by the Pacific Telephone and Telegraph Company for transmitting television signals from Los Angeles to Las Vegas, Calif.

The *TD-2* microwave system, which was developed by Bell Telephone Laboratories, was first placed in operation between New York City and Chicago, Ill., in September 1950. It operates in the frequency range 3,700–4,200 mc (megacycles) and is designed to handle from 1 to 12 broad-band channels, 6 in each direction of transmission along a route.

One of the frequency plans extensively used for the *TD-2* system is that shown in Fig. 1. As indicated, in a given direction from a repeater station the receiving frequencies are spaced by 80 mc and are centered between the transmitting frequencies which are also 80 mc apart. For branch routes intersecting at small angles, frequencies in between those shown are used to avoid interference. Ordinarily four antennas are employed at each repeater station, one pointed each way for transmitting and one each way for receiving. On some light routes a single antenna pointed each way may be used for both transmission and reception.

The *TD-2* system employs frequency modulation with a power output of only 1/2 watt for each broad-band channel. Two such broad-band channels, one for each direction of transmission, normally handle up to 600 telephone circuits or two television signals.

The route from Los Angeles to Las Vegas on which the radar interference was experienced was used for 1-way transmission on the frequencies 3,790 and 3,830 mc. Fig. 2 shows the location of this route. The interfering radar signals were entering the receiving antenna of the repeater at Strawberry Peak together with the desired signals on 3,700 mc. The radar signals were traced to spurious radiations of an *S*-band acquisition radar located in Los Angeles some 60 miles away.

As a result of this experience Bell Laboratories undertook an investigation of the causes of radar interference and means for avoiding or minimizing such interference. This effort was directed along the following lines:

1. Determination of the effect of radar-type emissions on the *TD-2* microwave system.
2. Study of the radiated frequency spectrum of certain high-power radars in use or planned for use by the military.
3. Investigation of means of suppressing the spurious and harmonic radiations from the radar emissions.

Effect of Radar-Type Signals on the TD-2 System

The first of these studies was carried out in the Laboratories by injecting varying known amounts of pulsed radar-type signals into the crystal mixer in the head end of a *TD-2* receiver, together with a simulated multichannel telephone signal corresponding to that normally received over an average microwave path. The results of these tests are illustrated in Fig. 3, where it is seen that as the peak value of the pulsed radar signal approaches the peak value of the desired signal the interfering telephone circuit noise rises very rapidly from negligible amounts for carrier-to-interference ratios of 5 db (decibels) or more, to intolerable amounts for carrier-to-interference ratios of −5 db or less. For the case of television, it was found that interference would result from radar signals 25 db weaker than those which cause interference to multiplex telephone circuits. The interference level for circuits carrying SAGE data signals is estimated to be about the same as for telephone. Although the tests just referred to were conducted on a *TD-2* receiver, field tests with other types of

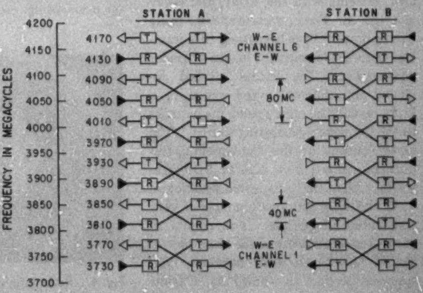


Fig. 1. TD-2 radio systems channel frequency arrangement

Paper 58-817, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N.Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 11, 1958.

R. D. Campbell is with the American Telephone and Telegraph Company, New York, N.Y.

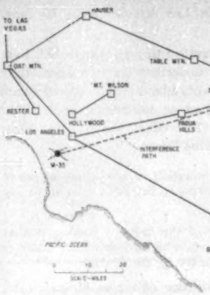


Fig. 2 (left). M-33 radar and microwave routes, Los Angeles area

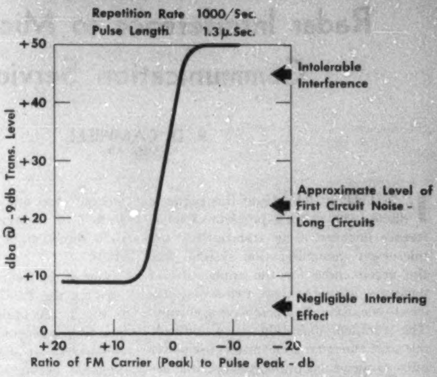


Fig. 3 (right). Effect of radio-frequency pulse interference on the TD-2 system

microwave equipment indicate similar effects.

Fig. 4 shows the effect of radar interference on a television picture. The separation between the dashes of interference seen in this picture depends on the radar repetition rate, and the length of each dash depends on the radar pulse length. In a telephone circuit the interference is evidenced by a rough buzzing noise.

Extra Band Radiation Characteristics of Radars

The Bell Laboratories investigation of radar frequency spectrum characteristics was carried out with the aid of a Polarad receiver with four radio-frequency heads mounted in a truck, with calibrated horns on the roof. The receiving equipment was capable of adjustment to any frequency in the range 2,100 to 11,260 mc. Fig. 5 shows the test truck. Attenuators were provided to prevent overloading the receiver with strong radar signals. The truck was also equipped with an oscilloscope to observe the signals being measured and a pulse generator which was used to simulate the radar signals and determine their amplitude by the substitution method. Through co-operative arrangements with the Army, the Navy and the Air Force, measurements using this equipment have been made on a number of radars commonly in use or under construction. The radars studied included nine different types, one using a klystron and the others using magnetrons as the frequency-generating source. The power output at the fundamental frequencies ranged from 0.14 to 5.0 megawatts. All of the units tested were found to be radiating spurious products of very substantial magnitude. The results of these tests are shown in the mass plot of Fig. 6, on which each harmonic and spurious product is shown as a point indicating its magnitude or power as ordinate versus the ratio of its measured frequency to the fundamental frequency of the radar as abscissa. As indicated, the spurious emissions range up to as much as 93 dbm (decibels referred to 1 milliwatt) effective radiated power, while many of the products observed had powers ranging from 40 to 70 dbm. The term "effective radiated power" (ERP) is used to express the peak power radiated by the radar including the gain of the radar antenna at the frequency in question. By way of comparison, it is interesting to note that the *TD-2* system has an ERP of only 66 dbm (0.5 watt power with a 39-db antenna gain). The radar harmonics shift up and down in frequency as the fundamental is changed. Likewise, some of the spurious outputs vary with magnetron tuning; others are fixed, or they vary only slightly. The latter probably depend on some geometry of the magnetron. All of the radars tested emitted extra band products which fell into one or more of the frequency bands used for common carrier services as well as in other bands. When these products fall directly on the communication frequencies being used, receiver selectivity affords no protection and the only means of avoiding interference is by antenna directivity; shielding by hills, trees, buildings, or other obstructions; and by physical separation of the radars from the communication receivers. Fortunately, most *TD-2* stations employ well shielded horn reflector or delay lens antennas, the directive characteristics of which are shown in Fig. 7. While such antennas greatly minimize the possibilities of interference from radars not directly in line with the receiving path, they are not capable of completely protecting against spurious emissions of powerful radars even though located "off-beam" at considerable distances.

Magnitude of Problem

Based on the previously mentioned measurements of the radar radiation

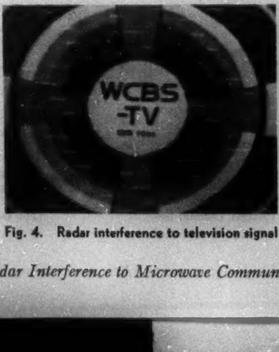


Fig. 4. Radar interference to television signal

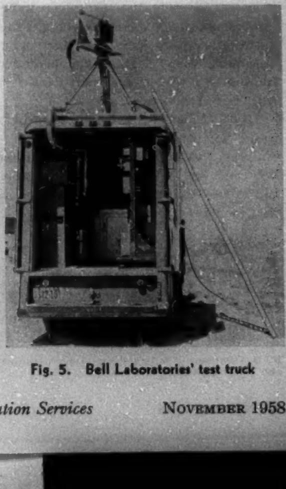


Fig. 5. Bell Laboratories' test truck

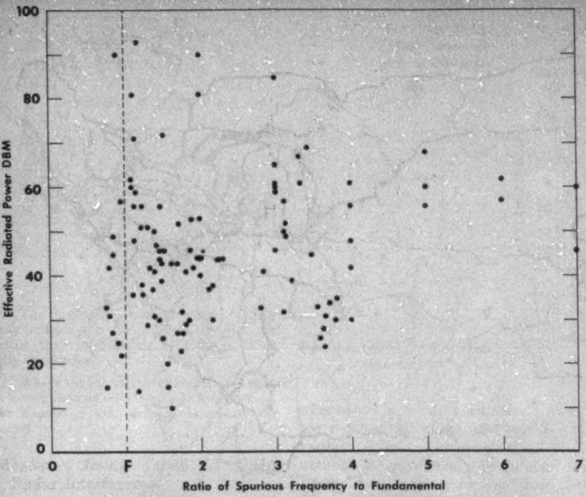


Fig. 6. Distribution of measured spurious and harmonic products of nine radars

characteristics, and their interfering effect on the *TD-2* system, it is possible to estimate whether interference will result from a radar at any specific location. Fig. 8 gives a graphical illustration of the extent of the interference areas that may be involved. This is a polar diagram centered on a *TD-2* station receiving signals over a nominal 30-mile path from a *TD-2* transmitter at point *T*. A radar transmitter emitting a spurious signal of 45-dbm ERP on a receiving channel being used for television reception may cause interference if it is located within the shaded area. The same area also applies to the case of a radar transmitter emitting a spurious signal of 70-dbm ERP on a receiving channel being used for telephone operations. As indicated in Fig. 6 many of the radars have spurious products as big or bigger than those assumed. The shaded area was calculated on the assumption that line-of-sight transmission exists for distances up to 30 miles with beyond-the-horizon smooth-earth transmission losses applying at greater distances. Fading of the desired signal of 30 db was also used in these calculations, since the *TD-2* system is designed to provide service in the presence of fades of this magnitude.

Considering the very appreciable size of the interference area indicated by Fig. 8, and the extensive use being made of microwave radio to provide the nation's communication needs, the reason for concern about the radar interference problem becomes apparent. Of all the microwave communication systems in service today, the Bell System alone operates more than 30,000 route miles consisting of some 1,300 individual stations scattered throughout the country. These facilities are used to provide 62,000 miles of intercity television circuits and about 13 million miles of long-distance telephone circuits. These circuits, which have been installed in a period of only 8 years comprise 80% of all the intercity video and about 25% of all long-distance telephone circuit mileage in the Bell System plant.

In addition to providing long-distance telephone and television services, these circuits are used for important industrial and governmental private-line services. The bulk of these circuits are provided

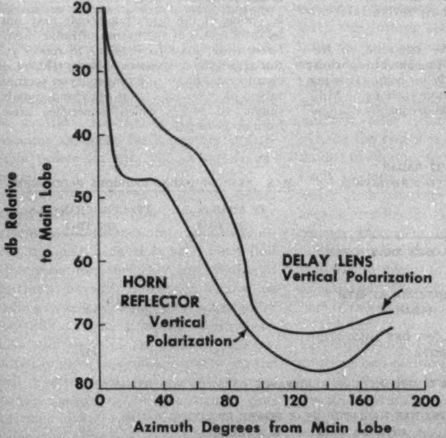


Fig. 7. Directive patterns of horn reflector and delay lens antennas

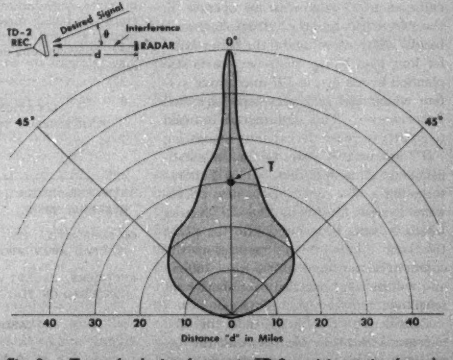


Fig. 8. Zone of radar interference to TD-2 receiving station for radar spurious of 45-dbm ERP on television channel or 70-dbm ERP on telephone channel

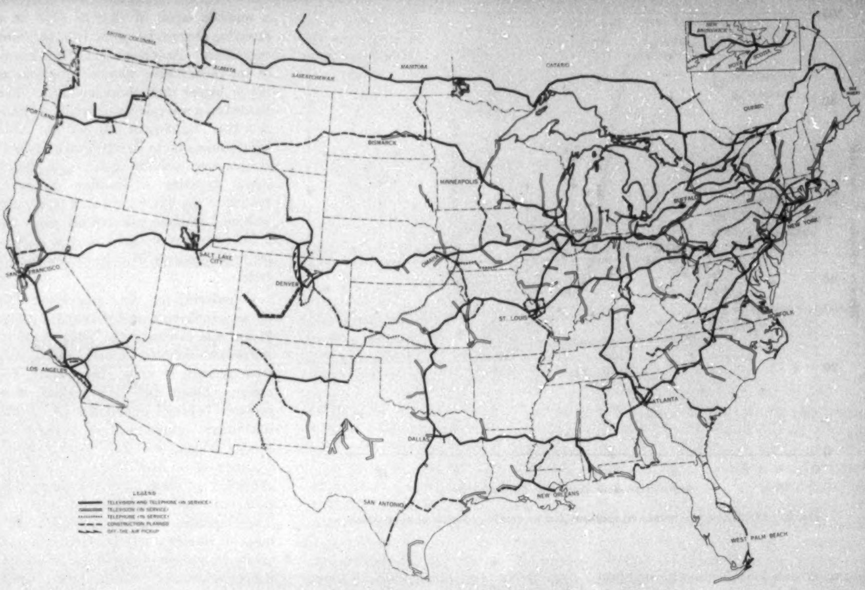

BRITISH COLUMBIA
ALBERTA
SASKATCHEWAN
MANITOBA
ONTARIO
QUEBEC
NEW BRUNSWICK
NOVA SCOTIA
PORTLAND
BISMARCK
MINNEAPOLIS
BUFFALO
NEW YORK
CHICAGO
OMAHA
SALT LAKE CITY
SAN FRANCISCO
DENVER
ST. LOUIS
NORFOLK
LOS ANGELES
ATLANTA
DALLAS
NEW ORLEANS
SAN ANTONIO
WEST PALM BEACH
LEGEND
TELEVISION AND TELEPHONE (IN SERVICE)
TELEVISION (IN SERVICE)
TELEPHONE (IN SERVICE)
CONSTRUCTION PLANNED
OFF-THE-AIR PICKUP

Fig. 9. Broad-band microwave radio relay systems: Bell and connecting routes

by broad-band microwave facilities, mostly *TD-2*, which operates in the 4,000-mc common-carrier frequency band. Fig. 9 shows the widespread extent of this broad-band microwave network. In addition to the routes shown here, the Bell Companies operate a large number of short-haul narrow-band microwave systems carrying up to a few dozen circuits each. These systems operate in the 890- to 940- and the 5,925- to 6,425-mc bands. Extensive use of the latter band for long-haul heavy route service is also planned by use of the *TH* microwave system which had been developed by Bell Laboratories. This system can be used to greatly increase the capacity of existing *TD-2* microwave routes or for the establishment of new routes. The Laboratories have also developed a new microwave system known as the *TJ* system which is being introduced into the plant this year. This system operates in the common carrier band 10,700 to 11,700 mc and will be used primarily for short-haul telephone and television services.

Considering the magnitude of the microwave network facilities and the extensive use being made of radar it is surprising that only a relatively small number of interference cases due to spurious radar emissions have been reported to date. This is believed due to the following principal factors:

1. In many critical locations the high-powered radars which are planned for use have not yet been placed in operation. As they go into service many severe cases are expected.

2. The radar may be operated on frequencies which have not resulted in spurious emissions falling directly on channels being used at exposed microwave stations. (Military considerations make frequency changes desirable.)

3. Many of the microwave receiving sites are shielded from the radars by hills, trees, or man-made obstructions.

4. On lightly loaded routes the receiving frequencies used may be selected so as to avoid the unwanted radar products. This is not possible on heavy routes.

5. Some of the operating people may not be fully aware of the nature of radar interference and may have failed to report it. Radar interference causes noise on telephone circuits, and pigeons (flying spots on picture tubes) or other effects on television signals similar to effects which sometimes arise from other causes.

LOCATION OF RADAR WITH RESPECT TO TD-2 RECEIVERS	MAX. ERP* OF RADAR SPURIOUS PROD.-DBM	
	TV CHANNELS ON TD-2	TELEPH. CHANNELS ON TD-2
ANYWHERE, EXCEPT CLOSER THAN A FEW HUNDRED YARDS	-35	-10
ON SAME SITE - TD-2 ANTENNA AT SITE POINTED AWAY FROM RADAR	+ 5	+30
ANYWHERE, EXCEPT ALONG ONE MILE STRIP CENTERED ON TD-2 ROUTE	+15	+40

(ABOVE LIMITS INCLUDE ALLOWANCE OF 30 DB FOR FADING)

* EFFECTIVE RADIATED PEAK POWER INCLUDING RADAR ANTENNA GAIN AT TD-2 FREQUENCY.

Fig. 10. Estimated tolerable levels of spurious radar products

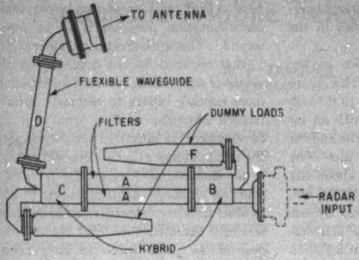


Fig. 11 (left). Filter assembly for M-33 radar

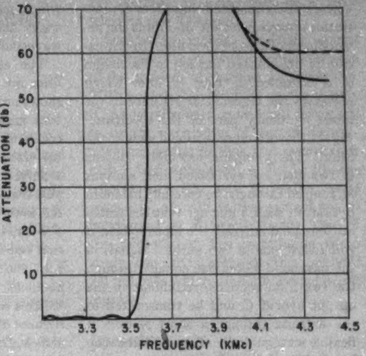


Fig. 12 (right). Filter characteristic

6. The radar or the communication facility may be used only part of the time, with the result that interference is avoided or is only intermittent.

7. Where the interference level is such that it is troublesome only during fading periods, the chance of identification is greatly reduced.

Measures Being Taken to Avoid Radar Interference

The question naturally arises as to whether radar interference to microwave systems might be minimized by increasing the power used by the communication systems. Consideration of this possibility indicates that it would not provide a satisfactory solution. The receiver sensitivities used are generally of the same order of magnitude as those used by radar receivers. To overcome interference from severe exposures to radar spurious emissions would require a terrific increase in the power of present microwave transmitters amounting to about 100 db. Not only would the use of powers such as this be wasteful, and technically and economically impracticable, but it would greatly aggravate the interference problem generally and would be contrary to the established radio principle of using the minimum practicable power for the type of service required.

Fortunately, the government agencies which use radars have recognized the seriousness of the radar interference problem, not only because it may affect Bell System services adversely, but because of its threat to other important microwave radio services including those operated by the military establishments themselves. This awareness of the problem has resulted in close co-ordination between the Bell System and the various governmental agencies in the selection of sites for new microwave communication and radar stations so as to minimize the possibilities of interference.

The problem is particularly severe where microwave communication facilities are required to serve radar sites. Advantage is taken wherever possible of natural or man-made obstructions to shield the microwave receiver from the direct radiation of the radars. Careful consideration is given to the frequencies used in order to avoid, as far as possible, those on which harmonic or other spurious products of the radars may fall. Where such measures fail to provide the necessary protection, it may be necessary to try special shielding of the communication receiving antennas, limitation of the frequency tuning ranges of the radars, or the use of sector blanking of the radars to prevent transmission directly toward the communication receiver.

It is recognized that such measures can be only temporary expedients at best. With the growing use in this country of both microwave communication services and radars, coupled with the trend toward higher powers for radar transmitters, it is obvious that means must be found to suppress radar radiations on frequencies outside the bands in which they are intended to operate, to values which are not likely to cause interference to other services. Such measures will not only give greater freedom in the selection of interference free sites for radar and communication facilities, but will also minimize the problems of mutual interference arising from the operation of large numbers of radars in given geographical areas.

The fact is recognized that it is probably impossible completely to eliminate spurious and harmonic products from the output of radar transmitters, and the question naturally arises as to how much suppression it is necessary to achieve to keep the interfering potentialities of the radars within reasonable bounds. The answer depends, of course, upon the characteristics of the facilities it is desired to protect as well as the power of the radar. Considering the problem of interference to the *TD-2* system, a table, Fig. 10, has been prepared, indicating the maximum effective radiated power that can be tolerated under various conditions from harmonic or spurious radar products falling in the receiving channel of a *TD-2* station. These figures assume an average 30-mile spacing of *TD-2* stations and include an allowance of 30 db for fading. From this table, it is seen that, to allow maximum freedom in the choice of location of radar stations and microwave communication sites, the emission of the radar on frequencies used for communication services should be limited to −35-dbm ERP. As indicated, somewhat greater levels can be tolerated where geographical separation and antenna discrimination permit. To achieve the −35-dbm levels with radars currently in use having spurious emissions of 40- to 70-dbm, ERP, a reduction of some 75 to 105 db is necessary.

Means of Suppressing Spurious Radar Emissions

Recognizing the seriousness of radar interference, the military services have initiated studies in their own organizations and through contracts with manufacturing, engineering, and development organizations seeking a solution to this problem. As part of the effort, Bell Laboratories has undertaken studies of means of suppressing the unwanted radiations from the *M-33* radar. In this work consideration was given to the possibility of suppressing the spurious signals within the magnetron, to the use of pulse-shaping techniques and to the use of a high-power filter in the transmitter output.

These studies indicated that the most practicable approach for the *M-33* was the development of a high-power filter. Investigation of this method has led to the design of a nonreflecting type of filter which preserves the transmitter stability in the mode frequency part of the spectrum.[1] Fig. 11 shows a cross-sectional view of the filter. The complete assembly consists of two identical evacuated filter sections *A*, coupled through two carefully balanced hybrids in such a manner that the transmitted signal entering the input hybrid *B* will follow one of two paths. Signals in the pass band will divide equally through the two filter sections, recombine in the output hybrid *C*, and be transmitted to the antenna through a short section of flexible waveguide *D*. Signals in the stop band will be reflected at the filters, recombined in the input hybrid and be dissipated in the input dummy load *F*. The output dummy load will absorb no power with a perfectly matched system, but it is required to handle reflections and mismatches in the output end of the filter which result from manufacturing tolerances and possible malfunctions in the waveguide assembly.

The electrical characteristics of the filter are shown in Fig. 12. This figure shows an insertion loss of 0.50 to 0.75 db over most of the pass band with an increase to just under 1.0 db at the ends of the band at 3,100 and 3,500 mc. The attenuation increases sharply above the pass band to more than 70 db at 3,700 mc, the lower edge of the band used by the *TD-2* radio relay system. The attenuation remains above 70 db from 3,700 to 3,950 mc and then falls off to a plateau of about 65 db up to at least 4,500 mc. While a single filter of this design will not reduce the *M-33* spurious products sufficiently to meet the −35 objective fully, it will greatly reduce the likelihood of interference from this type of radar.

Conclusions

Continued work on the radar interference problem is necessary if the usefulness of the radio spectrum for general communication purposes is to be preserved. This work divides into two parts: 1. that required for the development of means of suppressing unwanted emissions from existing radars to eliminate interference situations as they arise, and 2. development of improved techniques for the design of new radars which will result in the elimination of, or greatly reduced, spurious products at the source. Just as other interference problems which have seriously threatened to limit the usefulness of the radio spectrum have been overcome, so there is confidence that engineering techniques will be developed to conquer the radar interference problem.

Reference

1. EVACUATED WAVEGUIDE FILTER FOR SUPPRESSING SPURIOUS TRANSMISSION FROM HIGH-POWER S-BAND RADAR, **Harold A. Wheeler, Henry L. Bachman.** *Transactions*, Professional Group on Microwave Theory and Techniques Institute of Radio Engineers, New York, N. Y., vol. MTT-7, no. 1, Jan. 1959.

◆

Discussion

Harry R. Meahl (General Electric Company, Schenectady, N. Y.): The data Mr. Campbell has presented should prove useful to all who are interested in personnel safety as well as to communications engineers. Fig. 6, which shows the distribution of measured spurious and harmonic products of nine radars, reveals clearly why both broad-band and tuned monitors are needed at high-powered radar sites.

Some have assumed that both the near and far field patterns of radars may be calculated with enough accuracy for satisfactory application to personnel safety problems.

This article makes it obvious that both spurious and harmonic radiations can be too large to be ignored with impunity. The author is to be commended for achieving brevity together with broad coverage of his subject.

R. D. Campbell: I am very appreciative of the kind remarks made by Mr. Meahl about my paper on radar interference. Fig. 6 does show the need for broad-band and tuned monitors if those responsible for the operation of radars built to present-day standards are to keep track of their radiation characteristics. The indications from such monitors might be used as the basis for making adjustments of the radar equipment to minimize certain interfering products. However, such adjustments would become complicated and impracticable, particularly where a number of microwave receivers are being operated on different frequencies at exposed locations.

It seems obvious that the best solution is that suggested in the paper; i.e., the development of techniques to clean up the radar outputs so that they will not radiate appreciable amounts of energy on harmonic or other spurious frequencies. The output filter described in the paper illustrates one way in which this can be done. If such measures are applied to all radars, the broad-band and tuned monitors would not be needed, the interference problems would be minimized or eliminated, and there would be no hazard to personnel from such products. The personnel hazard from the radar fundamental would still be a major problem, however, as the fundamental is radiated at much higher energy levels than the spurious products.

Automatic Mobile Telephone System

R. P. DIMMER
NONMEMBER AIEE

MOBILE radio-telephone service has been available to the American public since its inauguration in St. Louis in 1946. While this service is quite extensively in use today by mobile subscribers, and to some extent in rural areas, the present system requires an operator to complete the connection between a mobile and land subscriber. This paper covers a development of equipment that will work satisfactorily into present dial telephone equipment. The system has been developed for the telephone section of the Rural Electrification Administration, and was the result of work done in the laboratory since 1946. The pilot system originally was installed in the North Florida telephone company's property in Branford, Fla., and recently was transferred to the same company's exchange in Live Oak, Fla.

It is anticipated that this new equipment will serve rural areas in three ways:

1. By providing telephone service to isolated ranches and farms, island locations, and rural areas that cannot be served by wire lines. This would be classified as "fixed mobile service."
2. By providing mobile subscriber service where it is desired for farm, commercial, or industrial use in automobiles, trucks, boats, and planes.
3. By providing radio-telephone communication between the telephone office and its own mobile units for maintenance and operation.

Some unusual problems are encounted when furnishing automatic dial telephone service to mobile subscribers. When a subscriber makes a conventional call, a d-c loop circuit is completed between his telephone and the central office equipment. This circuit causes a train of events to take place at the telephone equipment which takes over or seizes an idle incoming selector. As long as the subscriber's handset is in an off-hook condition, this circuit is held. When the subscriber dials a number, the circuit is advanced through a network of switches and relays, and eventually to the called party line. In a mobile radio system, a similar loop circuit cannot be maintained because the mobile radio transmitter does not remain operated except when the subscriber is talking. The operating mobile radio transmitter would correspond to the d-c loop circuit of the land subscriber. Consequently, seizure and holding of a line by the mobile subscriber must be locked in for a period of time (2.5 minutes has been found most satisfactory). It is also desirable to permit the mobile subscriber to release this seizure at any time prior to the end of this locked-in period. If his release signal is not received because atmospherics, for example, then the central office control circuits should be automatically released by the timer. If a connection is completed, it is also desirable to remove the release timer from the circuit until at least one party hangs up.

Essentially, this new system permits a regular telephone subscriber to dial any mobile subscriber in a community, or a mobile subscriber can dial any telephone subscriber without the aid of an operator except presently for toll calls. When long-distance toll dialing is in full use, it is expected that this type of service could also blend into this system. It also permits any mobile subscriber to call any other mobile subscriber through the control equipment using the existing telephone reverting call technique or a newer simplified method, which will be described later in this paper.

The primary consideration in the design of the equipment for this system was to make the procedure of making calls as nearly identical to conventional party line land service as possible. The one exception to this plan is the necessity of using a push-to-talk switch at the mobile end. The fixed mobile subscriber would not require this switch, because in this case the transmitting and receiving antennas can be separated by sufficient distance to prevent interaction.

The control equipment between the radio and telephone equipment was designed to connect to any type of automatic telephone line equipment with a conventional party line of a community automatic exchange (CAX) with four wires (two line wires, and two control wires). The present design allows for 20 mobile subscribers on a fully selective basis. The design can be expanded to any number of subscribers on a fully selective ring if desired.

Dial-to-dial mobile telephone service has been kept in obscurity partially because of the problem of charging for such calls. With no operator involved on these calls, it has usually been considered necessary to utilize some form of automatic toll-ticketing device to time and record such calls. Such equipment might be too expensive on a line with 20 parties; consequently, all calls in this system are to be charged on a flat-rate basis. One of the purposes of this installation is to study the advisability of this method. All toll calls will go through a regular toll operator.

Operation

In a mobile telephone system with a dial at both ends, signalling can be done generally in one of two ways: First, the dial can make or break a tone which modulates a radio carrier or, second, the dialed digit can be stored and translated into one or more steady tones, each tone or series of tones different for each digit from 1 to 0. In this system both methods are used. Signalling from the base station to the mobile unit is accomplished by sending out two tones (between 300 and 500 cycles) to operate a pair of resonant reed relays at the mobile end. This method is the most economical for the small number of mobile stations involved because the telephone equipment effectively translates the dialed number directly into an operation of one of ten sensing relays in the control equipment.

Signalling in the reverse direction (mobile to base station) is accomplished by transmitting a gate frequency (2,000 cycles) when the dial is taken off normal, and a pulse frequency (1,500 cycles) when the dial is released. These tones are received at the base station and are converted directly into d-c pulses to select the called land subscriber. This eliminates a larger number of oscillators at the mobile station, together with storage and translation equipment to change the digit dialed into multiple tones A simplified diagram of both signalling circuits is shown in Figs. 1 and 2.

In a call initialed by a land subscriber his radio transmitter is operated when the mobile radio number is dialed (usually on the second to the last digit dialed). Reception of this signal at all mobile units on this channel will operate their busy lamps. This indicates to the mobile subscriber that the channel is in use, but does not prevent him from breaking into an existing conversation if an

---

Paper 57-1020, recommended by the AIEE Communications Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted July 26, 1957; made available for printing March 21, 1958.

R. P. Dimmer is with the General Telephone Laboratories, Inc., Northlake, Ill.

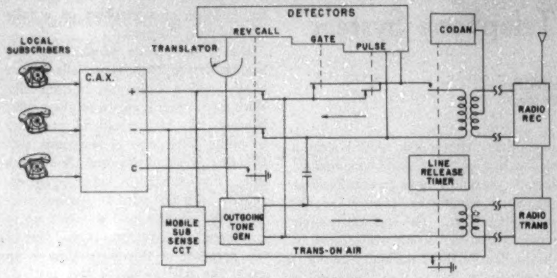


Fig. 1. Block diagram of land station

emergency exists. If he should break into an existing call, he cannot release the circuit on hang-up.

The called ringing voltage, usually obtained on the last digit dialed, will operate one of the ten sensing relays in the control circuit. These relays are a resonant vibrating type of device which will respond to any one of the ringing frequencies between $16\frac{2}{3}$ and 66 cycles per second. They will select two oscillator frequencies corresponding to the number of the subscriber being called, and place them on the radio carrier. These tones will follow the ringing period.

In the mobile unit of the called subscriber, reception of these tones will operate two resonant relays which in turn will operate the buzzer. Operations of these various relays and buzzer will follow the ringing period of the telephone system.

When the mobile subscriber answers, removal of his handset from the dash unit automatically places his radio transmitter on the air for 2 seconds. This brief signal at the base station receiver causes tripping of the ringing cutoff relay in the telephone equipment, and removes the ringing voltage and tones from the circuit. The circuit is now complete, and conversation can ensue.

If the wire line telephone should disconnect first, the connector would release and start to put the called mobile telephone on line lockout. The auxiliary cutoff relay at the radio equipment would then release along with the regular central office cutoff relay. This would open the central office loop and turn off the fixed transmitter without waiting for the disconnect from the mobile subscriber to release the central office loop and radio channel. Should the mobile subscriber then disconnect after the connection had been released by the wire line telephone the fixed radio transmitter would have been turned off and the radio loop already opened, so that the mobile disconnect would not do anything at the fixed radio station.

In a mobile-initiated call removal of the handset from the dash unit again automatically places its radio transmitter, plus tone, on the air for a moment. Reception of this signal turns on the wire-line radio transmitter, and seizes an idle incoming selector in the telephone equipment. Dial tone is returned via this transmitter to the mobile receiver. The mobile subscriber may then dial in the usual way. All tones (busy tone, ring-back tone, etc.) normally associated with automatic telephone equipment can be heard by the mobile subscriber.

When the dial in the mobile unit is taken off normal it turns on the mobile radio transmitter and sends out a tone to operate the gate relay in the land station. This relay remains operated until the end of the pulsed digit because of its slow-release characteristic. Operation of the gate relay removes a shunt on the contacts of the pulse relay and permits it to function. When the dial is released, the pulse relay follows, and breaks the d-c loop circuit into the telephone equipment at a pulse rate. At the end of the pulsed digit the gate relay falls back to normal. In this technique, it is necessary that both a long gate and short pulse signals be received before an actual selection can occur at the telephone exchange. In a stand-by condition it would be necessary to operate the gate detector with a long period of noise of the proper frequency at the same time that the pulse detector operates to cause false operation of the telephone equipment.

In a reverting call (mobile to mobile) the incoming pulses from one mobile station are translated in the control circuit into two tones for selecting the other mobile subscriber.

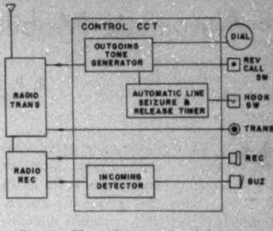


Fig. 2. Block diagram of mobile station

During the time that the mobile subscriber initiates the call to the answer by the wire line subscriber, the telephone line seizure is locked in for a 2.5-minute period. When the wire line subscriber answers, this timer is disconnected from the circuit, together with the gate and the reverting call detectors, to avoid operation on voice. This is accomplished by the use of a reverse battery relay on answer.

When the mobile subscriber hangs up, his radio transmitter is again placed on the air automatically, for 2 seconds with a long 1,500-cycle tone. Receipt of this disconnect signal opens the loop to the central office. If the mobile subscriber had been the first to disconnect, the receipt of this on-hook signal would be displayed back at the connector, so that the connection would be returned to normal when the calling party disconnects. If another mobile station

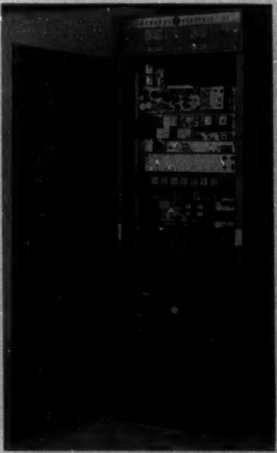


Fig. 3. Fixed or land station

Vol.
Pt. 1
195

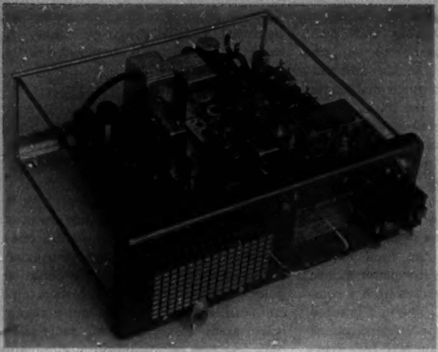


Fig. 4 (left). Mobile station

Fig. 6 (right). Fixed-mobile station desk unit

should attempt to call in the meantime the call would merely land on the already established connection and hence, the station would receive no dial tone. This corresponds to what happens on wire lines; it is necessary to hang up and wait for the called line to be freed by the original calling party.

To guard against failure to receive a disconnect signal from the calling party (the calling party does not restore the handset properly, or the line is in trouble) which would result in tying up the radio channel with the radio transmitter turned on, it is necessary that the telephone switching connectors serving the radio channel be equipped with connector time-out. The time-out equipment at the radio station is of no help on incoming calls to the mobile stations because the central office connector blocks off the calling party's supervisory signals, preventing them from being received at the radio station which is now on the called party's end of the connection.

With automatic time-out provided in the connector group serving the radio channel, any trouble condition resulting in failure to receive a disconnect signal from the operators office or interoffice trunk will be timed out, provided, of course, that the radio channel shows on-hook. In other words, trouble conditions on wire lines or on trunks cannot tie up the radio channel for more than the connector time-out period.

Fig. 5. Mobile station dash unit

If the mobile subscriber wishes to make a reverting call he would press the reverting call button on his dash unit for a moment. This places the mobile transmitter on the air with 2,500-cycle tone. At the base station, reception of this signal would BUSY out the radio circuit to any other land subscriber. This operation would also ready the reverting call translator. The mobile subscriber would now dial two digits: either 01 to 00, or 11 to 10. If he dials 10, for example, the first digit would select the ringing code, and the second digit would select the two ringing tones. When the second mobile subscriber answers, the tone generators and detectors would be removed from the circuit. Again the calling party has power to release the circuit. In this type of call one of the two mobile units must transmit for a moment every 2.5 minutes to prevent the systems' releasing on the timer.

Since automatic time-out is provided at the land radio station for outgoing calls, it is necessary for the mobile station to hang up after dialing an operator for a toll call, and be called back to eliminate this automatic time-out feature which would otherwise open the connection prematurely. This is because the operator trunks cannot reverse the calling party's battery supply; hence, if the mobile station attempted to hold the original connection, it would be automatically timed out and disconnected. The operator would advise the mobile station to hang up and then wait for the disconnect signal before attempting to dial back the station. If this particular disconnect signal should be lost, the operator would merely continue to hold the circuit until the radio time-out provided the disconnect signal automatically. It is also possible to eliminate this call-back requirement by having the mobile subscriber press his talk-to-listen switch momentarily every 2.5 minutes. This causes the automatic timer to reset to zero and start a new timing cycle. On calls of this type the gate and reverting call detectors are removed from operation after a 30-second delay to prevent their operation on speech.

In accordance with Federal Communications Commission requirements the call letters of the base station are transmitted at the end of every transmission. These are transmitted in continental code at a speed of 25 words per minute. One of the outgoing signalling oscillators (500 cycles per second) is used as the tone source.

Equipment

A typical base station is shown in Fig. 3. The unit comprises a 50-watt frequency-modulation transmitter, operating in the 152-megacycle band. These units occupy the upper part of the cabinet. The cabinet houses the control equipment in the lower section. On this panel are located the outgoing tone generators, the three incoming tone detectors, and the necessary relays, switches, and sensing devices to connect to the telephone equipment. A separate power supply, mounted on the bottom, furnishes both high voltage and 48 volts for relay operation, to permit separate operation of the control equipment if desired. For separate operation, four wires connect the control equipment to the telephone exchange, and four wires are required between the control equipment and the radio equipment. The control equipment is designed to be operated from a stronger, all-relay, or cross-bar type of telephone office.

Two types of mobile equipment were used in this trial installation, one 60-watt unit, and one 30-watt unit. Actually, the system will work with any mobile gear. The selector utilizes a 3-inch chassis so as to fit into a standard 17-inch case. It may be externally mounted if so desired. Fig. 4 shows a mobile unit with cover removed.

Fig. 5 shows the mobile dash unit furnished. This unit is of the cradle design, very similar to a conventional telephone except that it has a simple locking device to prevent the handset from bouncing out of the cradle on rough terrain. This unit contains an ON-OFF key-type switch on the side, and ON-OFF, BUSY, and ON-AIR lamps in the upper front. A small reverting call push button switch is mounted on the opposite side visible in Fig. 5. The dash unit contains a hinged bracket which enables the unit to be mounted on the top, front, or bottom of the dash panel of the car. A short cable extends from the dash unit to a small buzzer box usually mounted under this dash panel. This unit contains the volume control, squelch control, buzzer, and the necessary terminals to connect the radio equipment to the dash unit.

The equipment in the fixed mobile station is similar to the low-power mobile unit except that it operates from a commercial power source. The unit is rated in the 15-to-25 watt class. A desk phone is used in this application with the necessary indicating lamps as shown in Fig. 6. In place of a buzzer this unit uses a conventional 60-cycle gong-type ringer. A reverting call button is contained directly under the handset.

Conclusions

Since the development work on this project was started, a large amount of interest has been shown and many new uses considered for this type of a telephone system. There seem to be many places where it might be used to great advantage either as a way of offering quick temporary telephone service, or in places where telephone service does not warrent the expense of poles and wire. Where microwave circuits can eliminate outside plant between urban areas, so an automatic radio telephone system of this type can eliminate outside plant within the urban area.

Magnetic Field Strength in Oersteds Near Large Rectangular Conductors

J. H. MILLER
FELLOW AIEE

A HIGH VALUE of direct current in a conductor produces a high magnetic field which may affect electromagnetic instruments and control gear in the vicinity. The value of the magnetic field in oersteds at a distance d in feet, from the center of a large round conductor, solid or tubular, carrying a current of I amperes is given very simply by the equation,

$$H=0.00656\, I/d \tag{1}$$

A brief tabulation of some representative values is given in Table I.

In the electrochemical industries, including the manufacture of aluminum metal from alumina, the conductor is usually rectangular in shape, consisting of a group of rectangular bus bars. Since the conformation is not circular the preceding equation for field strength cannot be used directly.

To obtain a working procedure to determine the field strength adjacent to a rectangular conductor carrying a high current, it was decided that the laminated bus could be considered as a solid conductor within its outside outline dimensions and that for the purpose of this analysis the point at which the field was to be determined would be on an axis of the conductor cross section.

Referring to Fig. 1, dimensions a and b were assigned as shown, with d the distance from conductor center to the point P, at which the field was to be determined. To allow for a dimensionless formula, parameters were assigned where $A=a/b$, a sort of aspect ratio, and $D=d/b$. In collaboration with W. N. Goodwin, Jr., formerly Chief Engineer for the Weston Electrical Instrument Corporation, a complete integration over the entire rectangular current path for the field at point P gave the following:

$$H'=H\left[\frac{D}{A}\left\{(2D+A)\tan^{-1}\frac{1}{2D+A}-(2D-A)\tan^{-1}\frac{1}{2D-A}+\frac{1}{2}\log\frac{1+(2D+A)^2}{1+(2D-A)^2}\right\}\right] \tag{2}$$

In the foregoing, H is the field at point P if the conductor were circular in section and is obtained from equation 1. The remainder of the equation then may be considered as a modifying factor F, so that we might say $H'=HF$. In the long equation the angles corresponding to inverse tangents are in radians and natural logarithms are used.

Values of F have been calculated for rectangular conductors of various shapes, having a width at right angles to the axis to the point P in question from 20 times the thickness to a width 1/20th of the thickness, and these are plotted in the above chart, using however, the ratio $A=a/b$ to identify the several curves, and

Table I. Representative Values of Magnetic Field H, in Oersteds

	d, Feet					
I, Amperes	1	2	5	10	20	50
5,000	32.8	16.4	6.6	3.3	1.6	0.7
10,000	65.6	32.8	13.1	6.6	3.3	1.3
20,000	131.2	65.6	26.2	13.1	6.6	2.6
50,000	328.0	164.0	65.6	32.8	16.4	6.6
100,000	656.0	328.0	131.2	65.6	32.8	13.1

Paper 58-367, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 6, 1958; made available for printing April 1, 1958.

J. H. MILLER is with Weston Instruments, Division of Daystrom, Inc., Newark, N. J.

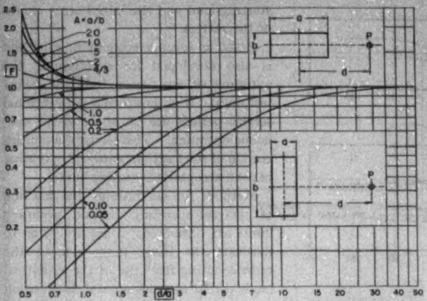


Fig. 1. Plot of values of modifying factor F, used in obtaining magnetic field strength in oersteds near a large rectangular conductor carrying a heavy current

d/a as the ordinates to simplify the presentation.

Essentially, then, to determine the flux density at point P, determine a/b and select the appropriate curve or interpolate between two curves. Determine d/a and select the appropriate vertical ordinate. The intersection of the ordinate and curve gives the value of F on the scale at the left. The value of the field in oersteds, is, then,

$$H' = 0.00656\, I/d(F) \qquad (3)$$

Note that a field of 5 oersteds is ordinarily the maximum field strength for which shielded instruments are rated. In higher fields additional shielding is required. Alternatively, self-shield structures such as the Weston core-magnet mechanism may be adequately shielded for upwards of one-hundred to several hundred oersteds depending on the particular mechanism and casing used.

◆

Discussion

Max Kramer (Aluminum Company of America, Cleveland, Ohio): The resistance paper analog method commonly used to determine magnetic flow lines set up by large rectangular bus structures will give errors caused by variations in the resistance constant of the paper, and errors inherent in the method, for flow lines close to the bus. It is possible by a certain procedure, to translate the flow lines into field strength, and the accuracy of this method is contingent on the trueness of the flow pattern.

The method of calculating magnetic field strength given in Mr. Miller's paper is sufficiently accurate and rapid to encourage common useage. The writer has checked this method against the point-source of current formula, $2I/R$, to determine field strengths of a bus having an outline dimension ratio of 4 to 1. The simple formula can be used with small error for points in space at a distance of at least ten times the bus smaller dimension, the error near the bus outline larger dimension being much less.

To avoid uncertainty in the calculated field strength for points in space, the method offered by Mr. Miller is now used exclusively by the writer. It is possible to calculate the field strengths for points on the two axes of a bus and to draw in the flux flow lines more accurately than can be done with the resistance analog. These flow lines can then be reproduced on the analog paper by ordinary methods of manipulation, and then ferrous structures can be "injected" into the setup to determine flow distortion and to re-evaluate the field strengths.

Some work along these lines has been done.

Evaluation of Single-Phase Metering Practices Based on Meter Performance and Load Characteristics

J. M. VANDERLECK
MEMBER AIEE

Synopsis: To evaluate such choices as rating, type, make, and service period for meters, a method is developed for determining the error in measurement of total energy taken by customers. Also, a method is developed for determining the optimum economical service period for meters. In a discussion of statistical methods for estimating the fraction of meters operating beyond acceptable accuracy limits, it is shown that normal distribution of meter errors should not be assumed. Findings are presented regarding meter errors, customers' load characteristics, error in total energy measurement, optimum service period, and fraction of meters operating beyond accuracy limits. It is concluded that the present ruling on length of service period should be changed to one that limits the fraction of meters beyond accuracy limits. This ruling should lead to improved metering and very significant savings because of longer service periods.

IN METERING the loads of single-phase customers, a large proportion of the cost is caused by the necessity for periodically servicing the watt-hour meters. The total cost of servicing a meter may well exceed the purchase price of the meter, even if the meter is serviced only three times during its life. The maximum length of the period between servicing operations (service period) is usually determined by a regulatory authority. Canadian Government regulations permit modern meters to operate 8 years before reverification of accuracy. No doubt the limit selected is based on judgment regarding the long-term accuracy of meters. It would be to the ultimate advantage of the utilities if the long-term accuracy were improved, because longer service periods would likely be permitted and, in addition, less revenue might be lost from the increase of minus errors as the length of service increases. It is estimated that roughly 3 million dollars a year is spent in Canada to service meters periodically, and roughly ten times this amount in the United States. This is based on an estimated average service period of 8 years, and an estimated average cost of 6 dollars per servicing operation. By reducing the need to service meters to one replacement by a new meter every 32 years, the servicing costs

---

Paper 58-425, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 1, 1958; made available for printing April 30, 1958.

J. M. Vanderleck is with The Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada.

The author acknowledges the valuable contributions on statistical methods and analysis by J. G. C. Templeton of The Hydro-Electric Power Commission of Ontario.

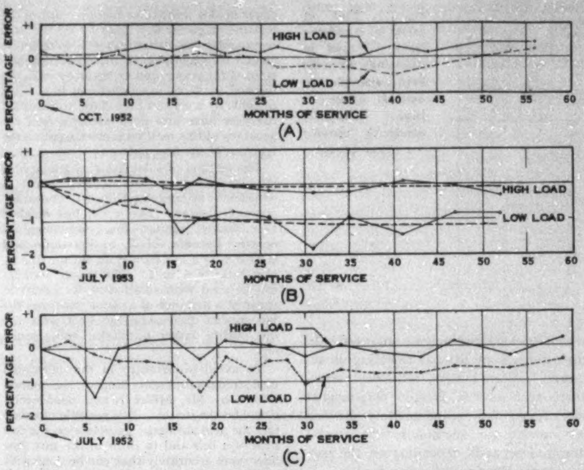


Fig. 1. Mean error of samples of meters operating outdoors. Five makes of meter in each sample with high load of 2,900 watts, low load of 460 watts, and 230-volt 3-wire 60-cycle 1-phase rating

A—Sample A: rating, 15 amperes nominal; 29 meters; continuous load, 2,900 watts; initially adjusted to zero error indoors

B—Sample B: rating, 25 amperes nominal; 15 meters; continuous load, 2,900 watts; initially adjusted to zero error indoors

C—Sample C: rating, 15 amperes nominal; 25 meters; continuous load, 460 watts; initially adjusted to zero error outdoors

become roughly one quarter (8/32) of the present-day costs, thereby saving millions of dollars.

The importance of knowing the error in measurement of the total energy taken by domestic and rural customers is reflected in the following figures. In 1955, the total revenue from such customers was about 200 million dollars in Canada[1] and about 3 billion dollars in the United States.[2] A loss of revenue of only 0.1% due to meter errors amounts to a loss of 3.2 million dollars per year to the electrical utilities in the two countries.

Information on the accuracy of meters is usually obtained from tests made at the end of a service period of several years. Such data are very useful to indicate the percentage of meters that have errors within specified limits at specified test points. However, from such data alone, one cannot determine the accuracy with which the total energy taken by customers has been measured. Consequently, the influence of meter errors on revenue is not known. For this reason, the author's utility has obtained additional test data on both meter performance and customers' energy-consumption (load) characteristics. The test results obtained from pilot samples of meters and customers are given in this paper. A method is developed for determining from the test results, the error in measurement of total energy. The method provides logical weighting of the meter errors found at the low-load and high-load test points, and adds the bias resulting from the minimum starting load and the initial error curves of the meters.

Based on revenue economics, a method is also developed for determining the optimum service period from the meter test data. It is shown that conventional service periods are far short of the economic optimum. This leads to a consideration of the fraction of meters operating beyond acceptable accuracy limits.

Methods for determining the fraction of meters operating beyond accuracy limits are compared. Data are given to illustrate the likely magnitude of the error resulting from use of the statistical method that assumes normal distribution of meter errors. The fraction of meters that measure the total consumption beyond limits is compared with the fraction that is found beyond limits from conventional as-found accuracy checks. Finally, consideration is given to the change of the fraction beyond limits relative to the length of service of the meters. Conclusions are reached with respect to the improvement that could result from a change in the regulations regarding the maximum service period.

It is suggested that the selection of meter types, ratings and operating practices should be based on the results of studies similar to those described in the paper.

Total Energy Measurement and Optimum Service Period

TESTS

To estimate the accuracy with which the total consumption would be measured by certain types or ratings of meters, test data were obtained on meters in simulated service for a number of years, and also on customers' load characteristics. The changes of error with duration of

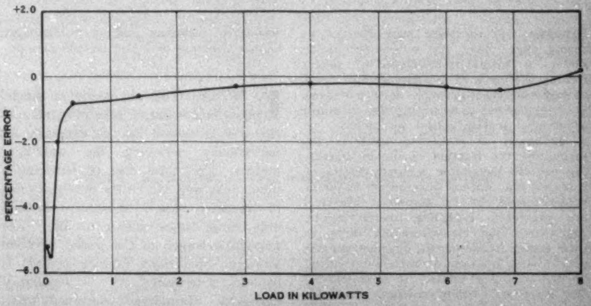


Fig. 2. Average error curve for sample B after 52 months of service; rating, 25 amperes, 230 volts, 3-wire, 60 cycles; 15 meters

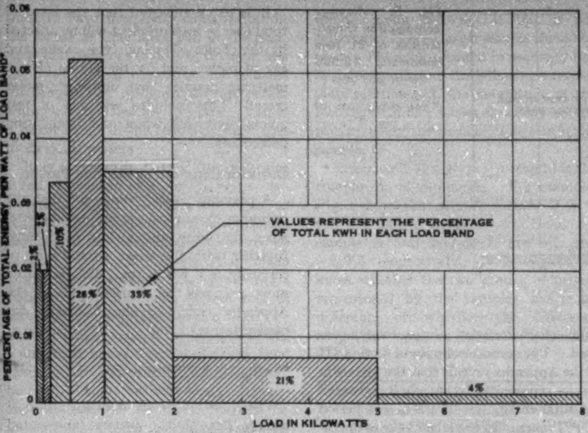


Fig. 3. Average consumption characteristics of 17 farm customers, 4,600 kw-hr per year

* If load band spans a range of 300 watts, then the percentage of total energy within the load band is 300 times the value of the ordinate

outdoor service are shown for three pilot samples of meters in Fig. 1. All meters were adjusted initially to substantially zero error, and then each meter was checked periodically for several years. Tests are continuing. It is apparent that the meters in sample A, Fig. 1(A), have maintained accuracy much better than those in sample B, Fig. 1(B). The errors of sample B meters at the high-load and low-load test points are used in the Appendix to compute the error in measurement of customers' total consumption over a period of several years. Also required to perform the aforementioned calculation are the average meter characteristics of error versus load, henceforth called the average error curve. In Fig. 2 is given the average error curve of sample B meters measured one day during the fifth year of service. Obviously, this curve does not apply at other times when the errors at the low-load and high-load test points are different from the values shown on the curve. The load curve at other times can be determined from a knowledge of the errors at the low-load and high-load test points. This subject is developed later in the paper.

Customers' load characteristics were obtained by conducting a pilot load survey. The object was to find the relative amounts of energy taken by customers in various load bands. This was accomplished by specially metering the load of each customer in the sample. Either an instantaneous recording wattmeter or a load-analyzing unit was installed at the customer's service entrance for an interval of 2 to 4 weeks. The survey extended throughout all seasons. The results are shown as histograms for three different groups of farm customers in Figs. 3 through 5. The groups differ in average annual consumption from 4,600 to 12,500 kw-hr (kilowatt-hour).

The histograms in Figs. 3 and 5 were obtained by integration of load charts. The charts were divided into 5-minute intervals, and the energy in each interval was allocated to the appropriate load band. Fig. 4 was obtained more simply from two load-analyzing meters which are described in the literature.[3] The meter used has four sets of dials that read directly the energy taken in each of four load subranges. The subranges for one meter are 0 to 240 watts, 240 to 600 watts, 600 to 2,400 watts, and all over 2,400 watts. The other meter enabled the 0 to 240-watt subrange to be divided into four, namely, 0 to 10 watts, 10 to 20 watts, 20 to 50 watts, and 50 to 240 watts.

Error of Total Energy Measurement

The errors of a meter at various loads and any point in time can be represented by an error curve. The curve may be expected to change with time in a complex manner. There may be both daily and seasonal fluctuations of each point on the curve, as well as a long-term drift.

In any event, if the mean value of error is determined at each load point during the entire interval of interest and if the percentage of total energy taken by the customer per increment of load is determined at each load point, then by combining the two curves, the error in measurement of the total consumption is obtained.

To obtain mean values of error at each load point during the entire interval it is necessary to measure periodically in service the meter errors at only a low-load test point and a high-load test point, and to measure the error curve of the meter only once. These measurements are sufficient because any change in the percentage error curve of a meter can be reduced approximately to a combination of a constant change independent of load, and a variable change inversely proportional to load. For example, when errors are small, a change in drag-magnet strength causes an equal change in percentage error at every load, and a change

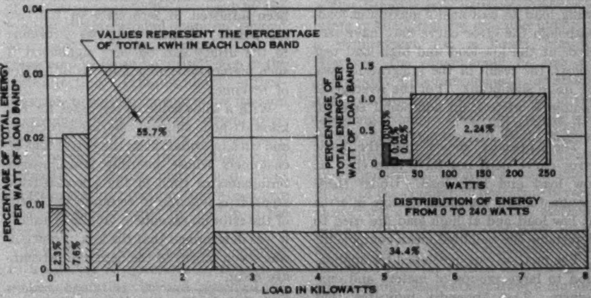


Fig. 4. Average consumption characteristics of 12 farm customers, 7,650 kw-hr per year

* See subcaption of Fig. 3

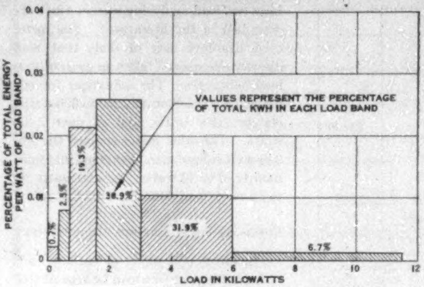


Fig. 5. Average consumption characteristics of 21 farm customers, 12,500 kw-hr per year

* See subcaption of Fig. 3

in friction has an effect inversely proportional to the load.

However, the method developed in section I of the Appendix avoids the calculation of the mean values of error at each load point. Instead, the error in total consumption is shown to be the sum of several components. One component is determined from the weighted mean of the mean values of error measured at low load and high load during the interval of interest. The weighting is based on the customer's load characteristics. A numerical example in section III of the Appendix reveals that the load characteristics given in Fig. 3 result in the following weighting when low load is 460 watts and high load is 2,900 watts:

$$(0.58\times\text{low-load error})+(0.42\times\text{high-load error})$$

Another error component results from the failure of the meter to register any of the energy taken below the minimum operating load. In the numerical example, this amounts to an error of about -0.1%. The remaining error component is the result of integrating the basic error curve of the meter with the customer's load characteristics from minimum operating load to customer's maximum load. Although the error curve may have zero error at the low-load and high-load test points, the result of the integration may be quite significant. For the average of sample B meters, an additional error of -0.2% was found.

The Appendix shows that the basic error curve need not have zero error at low load and high load. Under these circumstances, the mean change in error at low load and at high load are used to determine the weighted mean component.

The method was developed for application to large groups of meters and customers. The Appendix indicates that the method for one meter is applicable if average values of meter and customers' load characteristics, respectively, are used. The second example in section III of the Appendix reveals that the meters in service represented by sample B registered the total energy low by 0.9% in a period of $4^{1}/_{3}$ years. In the latest year of operation, Fig. 1(B) shows that sample B meter errors averaged -1.2% at low load and -0.15% at high load. The weighted mean error according to the aforementioned weighting is -0.8%. The error in measurement of total energy becomes -1.1% because of the other components of error.

When a utility at the end of a period of service depends on as-found tests in the meter shop to judge past performance, misleading results may ensue. For example, when sample B meters were removed from simulated service and shop-tested, the low-load and high-load test points were found to average -0.8% and -0.3% respectively. Neither these values nor the weighted mean values* would likely lead to a judgment that 1.1% of the total energy in the latest year of service had not been registered. Neither would it be expected that 0.9% of the total energy had not been registered in a period of $4^{1}/_{3}$ years, knowing that initially the meters were new and had been adjusted to zero error at the two test points. The resulting loss of revenue to the utility can be very significant if such meters determine millions of dollars of revenue.

With a knowledge of the rates for the kilowatt-hour steps in the rate structure, and with information on the percentage of customers whose monthly consumption terminates in each of the kilowatt-hour steps, it is possible to estimate the effect of the error on revenue. In one instance, it was determined that the loss of revenue was seven tenths of the negative percentage error.

It is to be expected that the error of total energy measurement will be affected by the choice of rating, type, make, and age of meter, and by the choice of any metering practice that influences meter errors. This includes selection of test points, and calibration and installation procedures.

Optimum Economical Service Period

A previous paper[3] points out that the optimum economical service period occurs when the mean annual loss of revenue together with the mean annual cost of servicing is a minimum. This is stated more precisely by equation 11 in the Appendix (refer to the Appendix for nomenclature):

$$kE_m+\frac{C}{t_1}=\text{a minimum} \qquad (11)$$

The question arises whether group B meters, i.e., all the meters represented by sample B, should be withdrawn from service, reconditioned, and reinstalled. A method based on equation 11 enables this question to be answered. However, the method presumes that after reconditioning, the meters will be returned to service and will repeat the original performance. The economics of replacing the meters by superior ones is not new, and therefore is not included in this paper. The performance of sample A and sample C meters as shown in Fig. 1 indicates that a different rating can reduce the loss of revenue substantially.

Equation 12 in the Appendix reveals that the optimum economical service period occurs at the time when the rate of loss of revenue less the mean annual loss of revenue up to this time reaches the mean annual cost of servicing up to this time. Whether the optimum service period has been reached can be checked with a knowledge of the errors at the low-load and high-load test points during the entire period of service. It is not necessary to know the average error curve (error versus load) of any meter or the whole group at any time. The method is developed in section II of the Appendix, and a numerical example is given for group B meters for $4^{1}/_{3}$ years of service in section IV. The example shows that with:

1. an assumed average annual revenue per customer of $80,
2. the percentage loss of revenue equal to seven tenths of the minus percentage error of total energy measurement, and
3. the cost of servicing equal to $6 per operation per meter,

then the optimum service period has not been reached. In fact, if the meters con-

---

* The Code for Electricity Meters[4] indicates two methods: 1. equal weighting, and 2. high-load error four times the weight of low-load error $(0.2\times\text{low-load error}+0.8\times\text{high-load error})$.

Vol.
Pt. 1
195

tinue to have increasing minus errors in the future as they have had in the past, such that the rate of loss of revenue less the mean annual loss of revenue remains unchanged, the optimum service period will not be reached for 48 years. This is a direct result of the relatively high cost of servicing meters.

It is noteworthy that if the in-service trend of a group of meters is first to accumulate negative errors, and then to stabilize at a constant error, an optimum economical service period may never be reached. This is because the mean annual loss of revenue kE_m approaches a constant, and the mean annual cost of servicing C/t_1 continues to diminish.

On the other hand, if the trend is toward a uniformly increasing negative error from the start, the optimum service period may be reached comparatively soon. A hypothetical example in the Appendix indicates an optimum period of 14.6 years, if the negative error in measurement of total energy increases 0.1% per year. It may be noted that these hypothetical meters have measured the total energy in 14.6 years with less error than group B meters have in $4^1/_3$ years.

It appears from the curves in Fig. 1 that the optimum economical service period of groups A and C meters will be longer than that of group B meters. Because such periods are greater than the normal lives of meters, it may be assumed tentatively that every endeavor should be made to lengthen the existing period to equal the life of the meter. This leads one to consider other limitations on the length of the service period, such as the fraction of meters operating beyond acceptable limits of error. This subject follows the next section which discusses the significance of results from small samples.

SIGNIFICANCE OF MEAN VALUES

Before drawing conclusions from the test results of samples as small as the pilot samples A, B, and C, the significance of the results should be determined. This can be accomplished readily by assuming that meter errors are distributed normally, and by using published probability methods.[5] For example, during the second and third years of operation of the 29 meters in sample A, the standard deviation of the errors was 1.1% assuming the variance of the group represented by the sample did not change during the interval. Based on these data, the 95% confidence limits for the mean of 29 observations are ±0.4%. That is, there is a 95% chance of being right when it is stated that the mean error for the group is equal to the sample mean ±0.4%. If this is not sufficiently close, a larger sample should be used. The improvement in accuracy can be determined from the following approximate relationship: the magnitude of the confidence limits varies inversely as the square root of the number of meters in the sample.

Reference 5 also gives a method for the comparison of two means. For example, Fig. 1 shows that after 50 months of service, the mean errors at low load for sample A and sample B are 0% and −0.9% respectively. It is desired to know whether the two groups of meters represented by the samples are really different, or whether the difference in the two samples occurred by chance. The method requires these data:

Sample	No. of Meters in Sample	Sample Mean	Standard Deviation
A	29	0	1.5
B	15	−0.9	1.9

The result of the computation indicates that group A meters are really different from group B meters with a 10% level of significance. That is, there is a 90% chance of being correct in the estimate that the mean error of group A meters is different from that of group B meters. Larger samples are required if more certainty about the difference is desired.

Although the sizes of samples A and B are small, it is believed that the inferences drawn from the test results are correct.

Fraction of Meters Beyond Error Limits

In at least one utility where selective sample testing has been adopted, the length of the service period for any group of meters appears to be governed by the fraction of meters found operating beyond acceptable accuracy limits.[7] This is considered to be an important step in the right direction. If the regulatory authorities were to specify the maximum fraction of meters allowed to operate beyond specified limits as the criterion for the maximum service period, then utilities could endeavor to choose meters and adopt meter operating practices that would give a service period equal to the life of the meter.

It would be desirable to have answers to a number of questions in respect to meters operating beyond specified accuracy limits. For example, what is an acceptable method for determining the fraction? Is it correct to assume Gaussian (normal) distribution of meter errors and thus take advantage of the method that employs the standard deviation? How large is the fraction for existing meters at the end of present service periods? How does this compare with the fraction that measures the total energy of customers beyond the limits? Does the fraction of meters that are too fast increase with the length of service? An attempt to answer some of these questions is made in the following paragraphs.

SAMPLE SIZE AND CONFIDENCE LIMITS

Several statistical methods are available for estimating from a sample of meters the fraction operating beyond specified accuracy limits. A discussion of two methods follows. The direct method gives an estimate equal to the fraction observed beyond limits in the sample. Another common method gives an estimate based on: 1. an assumption of normal distribution of the meter errors in the population that is represented by the sample, and 2. the values of the means and standard deviation of the observations of the sample.[7] The reliability of the results depends on the sample size in both methods, and on the accuracy of the assumption in the second method. The reliability of the results can be expressed in terms of confidence limits, as in the following example.

Suppose in a sample of 300 meters, drawn at random from a large population, 3% are observed directly to operate faster than a given accuracy limit. For this sample size, the 95% confidence limits are $1^1/_2$% and $5^1/_2$%.[8] This means there is a 95% chance of being correct in the statement that the fraction of meters in the large population operating faster than the accuracy limit is between $1^1/_2$% and $5^1/_2$%.

The broad range between confidence limits in the foregoing example indicates a requirement for a larger sample. However, the range could be reduced appreciably if the sample mean and standard deviation were used. Suppose in the foregoing example where the sample size is 300, that the method that assumes normality and employs the sample mean, and standard deviation gives an estimate of 3% of the population operating beyond the accuracy limit. Probability methods indicate that the 95% confidence limits are 2% and $4^1/_2$%. Unfortunately, the results of tests for normal distribution of the as-found observations from a number of groups of meters reveal that the observations were not distributed normally. Altogether the as-found checks made at

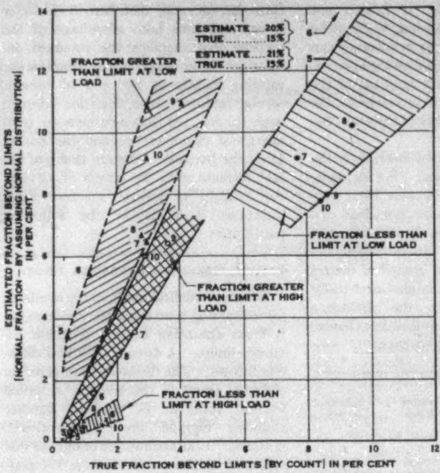


Fig. 6. Relationship between estimated and true fractions of meters beyond limits. Numerals beside plotted points identify the group of meters by years of service; no group contains less than 1,150 meters

low load and high load on five groups of meters were tested. The groups are identified in Fig. 6 by years of service of the meters, which ranged from 6 to 10 years. The number of meters in each group ranged from 1,309 to 6,214. The effects of the abnormal distributions are presented in the following.

ABNORMAL DISTRIBUTION OF METER ERRORS

An illustration of the typical errors that may result from assuming normal distribution is given in Fig. 6. For the five groups of meters mentioned in the foregoing, and one additional group of 1,150 meters, is shown the relationship between the estimated and true fractions of meters operating beyond limits of ±3% error. The estimated fraction was calculated from the mean error and standard deviation of the particular group, assuming normal distribution. The true fraction was determined by actual count of those meters beyond limits. Fig. 6 suggests that when the value of the fraction beyond limits is required, it should be assumed generally that the distribution is not normal unless proved otherwise. For brevity, the estimated fraction assuming normal distribution is termed the "normal" fraction. Note that the true fractions of meters below the −3% limit at low load, and the true fractions above the +3% limit at high load agree somewhat with the corresponding normal fractions. However, the true fractions above the limit at low load are substantially less than the normal fractions, and the true fractions below the limit at high load are substantially more than the normal fractions.

The observations on the group of 1,150 comprise as-found shop checks made on one rating of meter after 4 to 5 years of service. The meters were drawn from one municipality, in contrast to the other five groups which were drawn from various sources. In every instance, observations of meters that did not operate at either low load or high load were not processed.

These findings differ from those reported in a previous publication wherein a normal distribution was indicated for meters that had not been in service.[9] Evidently a reasonable interval of service destroys the normal distribution, at least with respect to the tails of the distribution curve.

Because normal distribution cannot be assumed, a larger sample size is required to obtain satisfactory results. When 95% confidence limits are desired for an estimate based on the fraction of observations found beyond an accuracy limit by direct counting in a sample, values can be obtained by reference to a published chart.[8] The chart gives confidence limits for sample sizes from 8 to 1,000, and for any fraction in the sample beyond a limit. The chart indicates that for 3% of a sample of 1,000 beyond an accuracy limit, the 95% confidence limits are 2% and 4%. Such results indicate that a sample of 1,000 meters will likely be acceptable.

ERROR IN METERING A YEAR'S CONSUMPTION

The accuracy of a meter is usually judged by the amount of its error at two or three selected test points. Because the function of the meter is to measure the total energy taken by a customer with reasonable accuracy, the best criterion for judgment is the error in measurement of total energy. A period of one year is suggested as a desirable length of time for observing the error. The difficulty of making such observations has precluded judgment on this basis in the past. However, the method outlined previously is practical and probably of sufficient accuracy. This method was used to determine the error in total energy measurement for each of many meters. From this was found the relation between the fraction of meters operating beyond accuracy limits at specified test points and the fraction that measured a year's consumption beyond the same limits. First, data are given for the group of 1,150 meters for which some results were reported heretofore; second, data are given for the meters in sample A.

Table I. Comparison of Fractions of Meters* Beyond Accuracy Limits for Conventional Checks and for Long Period of Operation

Condition	Fraction Slower Than −3% Error, Per Cent		Fraction Faster Than +3% Error, Per Cent		Total Fraction Beyond ±3% Error, Per Cent	
	Conventional Shop Checks	Long Period on Customers' Loads	Conventional Shop Checks	Long Period on Customers' Loads	Conventional Shop Checks	Long Period on Customers' Loads
(1)	(2)	(3)	(4)	(5)	(6)	(7)
High load	0.3		0.4		0.7	
Low load	14.9		0.4		15.3	
High load and low load combined†	15.0		0.5		15.5	
Customers' loads		6.3		0.3		6.6

* Sample of 1,150 meters (five makes) having 4 and 5 years of service, rated 25 amperes, 230 volts, 3 wire, 90 cycles.

† A meter is counted outside limits if found outside at either high load or low load.

Table II. Comparison of Fractions of Meters Beyond Accuracy Limits for Conventional Checks and for Long Period of Operation

		Percentage Error From Conventional In-Situ Checks		Percentage Error in Measurement for Long Period			Fraction Slower Than −2% Error, Per Cent			Fraction Faster Than +2% Error, Per Cent		
							Conventional In-Situ Checks			Conventional In-Situ Checks		
Period, Months	Test Point*	Mean†	Standard Deviation‡	Standard Deviation§	Mean‖	Standard Deviation‖	By Count	Assume Normality	Long Period Assume Normality	By Count¶	Assume Normality	Long Period Assume Normality
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
0 to 12	HL	+0.2	1.0	1.2	−0.1	0.9		$1^{1}/_{2}$	$1^{1}/_{2}$		$4^{1}/_{2}$	1
	LL	−0.1	1.5					9			8	
15 to 25	HL	+0.2	1.1	0.9	0	1.1		2	3		$5^{1}/_{2}$	3
	LL	+0.1	1.4					$6^{1}/_{2}$			$7^{1}/_{2}$	
27 to 40	HL	+0.2	1.1	1.0	−0.2	1.1		$2^{1}/_{2}$	$5^{1}/_{2}$		5	3
	LL	−0.3	1.4					11			$4^{1}/_{2}$	
44 to 56	HL	+0.3	1.3	1.1	0	1.2		2	5		$6^{1}/_{2}$	5
	LL	0	1.5					9			$8^{1}/_{2}$	
Average values	HL	+0 2			−0.1		$7^{1}/_{2}$	2	4	4	$5^{1}/_{2}$	3
	LL	−0.1						9			7	

* HL = high load, LL = low load.

† The mean in column 3 is the algebraic average of all observations made during the period.

‡ The standard deviation in column 4 is the square root of the mean of the variances determined for each set of 29 observations made during the period.

§ Column 5 is the standard deviation of the accumulated register readings during the period.

‖ Columns 6 and 7 pertain to the errors computed when measuring customers' loads.

¶ In column 11 a meter is counted outside the accuracy limit if found outside at either low load or high load.

Sample of 1,150 Meters. The sample of 1,150 meters resulted from excluding eight inoperative meters from a larger sample. None of the eight operated at low load, but five of them operated at high load. The results of as-found conventional shop checks are shown in Table I. By actual count, the fraction of meters found beyond the ±3% accuracy limits was 15.5%, which is comprised of 15% too slow and 0.5% too fast. However, after computing for each meter the error in measuring customers' loads, it was found that the fraction of slow meters was reduced to 6.3%, and the fast meters to 0.3%, resulting in a total of 6.6% operating beyond the ±3% accuracy limits. This is a very significant reduction from the former value, namely 15.5%.

The eight inoperative meters could have been included in the results. However, it is preferred to consider these separately. The three that did not operate at high load would normally accumulate no registration in service and thus be promptly discovered. The other five constitute a liability that would add about 0.5 to every value in columns 2, 3, 6, and 7 of Table I. This change does not affect the general conclusions drawn from Table I. To compute the error when measuring customers' loads, the following formula was used for each meter:

$$E = 0.4\ e_H + 0.6\ e_L - 0.1\% \text{ error}$$

The constants are those in equation 15 rounded-off. (Refer to section III of the Appendix.) It is assumed that customers' load characteristics are those given in Fig. 3, and that the as-found checks at high load and low load, namely e_H and e_L respectively, are the mean errors at these two loads in service. It is also assumed that the component of error that results from integrating the basic error curve of the meter with the customer's load characteristics is negligible. Some error has been introduced by the use of average instead of individual customers' load characteristics. The significance of this error remains to be investigated.

The results in Table I for this sample of meters demonstrate that the fraction found operating too slow from as-found shop checks is much greater than the actual fraction too slow in the measurement of energy taken over a long interval. However, the fraction found operating too fast is not much greater than the actual fraction too fast in the measurement of total energy.

Sample A, 29 Meters. Although the sample size is very small, results for sample A are given because this sample has been checked periodically during a 56-month period of simulated service outdoors. Consequently, the mean error of each meter at low load and high load over four intervals ranging from 10 to 13 months is well known. In addition, it is established that at one time all meters were correctly adjusted at low load and high load. This should yield greater accuracy in the determination of error in measurement of energy taken over a long period.

From the *in situ* checks, the fraction was determined of meters found operating beyond accuracy limits of ±2% error. The result is compared in Table II with the estimated fraction that would have operated beyond the same limits when measuring the energy taken by typical customers during approximately a year. Table II reveals that the fraction measuring a year's energy beyond limits is significantly less than the fraction found by *in situ* checks. A discussion of the contents of Table II follows.

During the course of 56 months, 18 sets of observations were made of the meters in sample A. Because of the small sample size, the sample mean and standard deviation likely varied more than the corresponding population mean and standard deviation. Consequently, for each of the four periods of operation in Table II, a common mean and variance were assumed for the observations. The error in measurement over each period of operation was computed by the method outlined in sections I and III of the Appendix. Column 7 in Table II indicates a lower variance for the error in measurement over a long period (one year) of operation. Consequently, columns 10 and 13 show the estimated fractions beyond limits for the long period to be less than the estimated fractions beyond limits based on conventional *in situ* checks. For example, the lowest section of Table II shows that on the average, it is estimated that $5^{1}/_{2}$% of the population of meters would be found faster than the limit at high load, and 7% faster at low load; i.e., between 7% and $12^{1}/_{2}$% of the meters would be faster than the limit. The estimate of the fraction fast for the long interval of measurements is 3%, which is less than half of the former figure. Although the individual estimates may be incorrect, it is believed that the ratios of the estimates are indicative of the truth. Column 11 shows that the actual average fraction of the sample faster than the limit was 4%. The actual average fraction of the sample for the long period of measurement was only $2\frac{1}{2}\%$, which supports the belief stated in the foregoing.

The observations of the register readings of the meters in sample A lend further support to the indication that appreciably fewer meters measure a year's total energy beyond specified limits than are found beyond limits from conventional *in situ* checks. All sample A meters were operating on the same load, namely 2,900 watts. Table II shows that for each of the four periods of operation except the first, the standard deviation of the accumulated register readings, column 5, is significantly less than the standard deviation of the conventional *in situ* checks at high load column 4.

The evidence from the foregoing two samples leads to the probability that far fewer meters, possibly one half or less, will measure customers' consumptions over a 1-year period beyond accuracy limits than will be found operating beyond limits when checked conventionally *in situ* or in the meter shop at two test points. However, the improvement may be mostly in the fraction of meters that are too slow.

Increase of Fraction With Time

Published data reveal that the fraction of meters that are too slow increases with length of service. However, the fraction too fast shows no definite tendency to increase with length of service. A review of some data follows.

Frequency distributions of meter errors after varying periods of service are given by Golds.[8] For one make of meter, the fraction above $+2\frac{1}{2}\%$ error at either low load or high load was substantially the same after 14 years of service as that after 2 years of service. For another make, comparing the data for 4 and 12 years of service, the fraction too fast at low load was substantially unchanged, but at high load increased from 4% to 6% approximately. For a third make, comparing the data for 2 and 14 years of service, the fraction too fast at low load increased roughly from 2% to 4%, but at high load fell from about 5% to $2\frac{1}{2}\%$.

Similar data published by the Canadian Electrical Association[10] for groups of meters after 6 to 10 years of service reveal no substantial increase in the fast fraction at low load, but possibly an increase in the fast fraction at high load. The sample sizes ranged from 1,830 to 16,000 meters approximately. The fraction faster than $+3\%$ error at low load was roughly 2%. The fraction faster than $+3\%$ error at high load increased from about $1\frac{1}{2}\%$ to $2\frac{1}{2}\%$.

The data in Fig. 6 indicate a trend toward an increase in the fraction too fast as the period of service lengthens. However, most of these data are not independent of the data in the previous reference.[10]

It is possible that none of the changes in the fraction too fast mentioned in the foregoing is statistically significant because of the sampling methods used. Where the same sample has been observed for several years, namely sample A, no trend has yet been indicated with respect to a change in the fraction too fast. This is shown in column 12 of Table II.

Optimum Service Period Based on Fraction Too Fast

It has been suggested that the overcharge for energy due to the fraction of meters that operate faster than the positive limit of error could serve as a basis for determining the optimum economical service period from the customer's viewpoint. However, unless the periodic servicing of meters reduces the overcharge, the basis does not appear sound. The possibility that the fraction too fast does not increase with time suggests that the overcharge does not increase with time. It follows that periodic servicing of meters would not reduce the overcharge under such circumstances.

The observations of the meters in sample A were analyzed to find whether the average error increased of the fraction having positive errors. For each meter was calculated the error in measurement of the energy taken by a customer during each of the four successive periods of operation shown in Table II. Positive errors resulted for about half of the meters. The fraction fast averaged $+0.6\%$ error for the first year of operation, and $+0.8\%$ error for the latest year. The increase does not appear to be significant to date.

Summary on Performance of Meters

The performance has been discussed of types of meters that are no longer being manufactured. All meters are now 5 or more years old. However, in some respects even these meters appear to behave better than is generally realized. Indications are as follows:

1. The fraction of meters that measure the total energy with an error beyond specified limits is less than the fraction found operating beyond the limits as a result of conventional as-found accuracy checks.
2. The fraction of meters that operate faster than the specified limit does not appear to increase significantly with the length of the period of service.
3. The overcharge for energy due to the fraction of meters that operate fast, or faster than the limit, does not appear to increase significantly with the length of service.
4. The optimum service period based on loss of revenue and cost of servicing probably is greater than the normal life of meters.

Under these circumstances, servicing meters every 6, 8, or 10 years may well be an unnecessary expenditure with no benefit to the utility or the customer, who must pay ultimately for the service through the charges made for electric energy.

Conclusions

In order to evaluate single-phase metering practices, a method has been developed for determining the error in measurement of total energy taken by customers. The method gives the weighting to be applied to measured errors at the low-load and high-load test points. To this error is added that resulting from the basic error curve of the meter, including the loss of registration of energy taken at loads below the minimum operating load of the meter. The results depend on the customers' load characteristics.

It has been shown that as-found checks in the meter shop may give misleading results regarding the magnitude of error of total energy measurement.

A method has been developed for determining the optimum service period based on revenue economics. Examples indicate that the optimum period likely exceeds the normal life of a meter, and thus it appears desirable to eliminate periodic servicing entirely. In this event, meters would be replaced only at the end of life, normally about 30 years.

After a few years of service, as-found checks generally do not have a Gaussian (normal) distribution. Consequently, the fraction of meters operating beyond acceptable accuracy limits should be found by statistical methods that are not based on this assumption. For reasonable accuracy, the sample size should be about 1,000 for any selected group of meters.

Error data of meters indicate that the

fraction of meters beyond acceptable accuracy limits is appreciably less when based on total energy measurement than when based on conventional as-found accuracy checks. The fraction of meters that operate faster than the upper limit does not appear to increase significantly with length of service. Neither does the overcharge for energy due to a fast group appear to increase significantly with length of service.

The utilities should welcome a change from the present ruling that specifies the maximum years of service of meters before reverification to a ruling that limits the fraction of meters operating beyond stated accuracy limits. This would encourage the replacement of inadequate meters by better ratings and types, and lead ultimately to the installation of meters for a life of say 30 years. The result should be a saving of millions of dollars to the utilities and their customers.

Appendix

Nomenclature

A = constant component of Δe
a = a constant, see equation 6
B = constant relating the variable component of Δe to P
b = constant increase in error per year
C = average cost of one servicing operation of one meter, dollars
c = a coefficient, see equation 7
E = percentage error in measurement of total energy taken by a customer over a period of one or more years
E_m = percentage error in measurement of total energy taken by a group of customers over a period of one or more years
e = percentage error of watt-hour meter at load P and time t
e' = percentage error of watt-hour meter at load P and time t'
e_H = percentage error of watt-hour meter at load P_H and time t
e_H' = percentage error of watt-hour meter at load P_H and time t'
e_{H_1} = average percentage error of group of watt-hour meters at load P_H and time t_1
$[e_H]_{t_0}^{t_1}$ = mean value of e_H

$$=\frac{1}{t_1-t_0}\int_{t_0}^{t_1} e_H\times dt$$

e_L = percentage error of watthour meter at load P_L and time t
e_L' = percentage error of watt-hour meter at load P_L and time t'
e_{L_1} = average percentage error of group of watt-hour meters at load P_L and time t_1
$[e_L]_{t_0}^{t_1}$ = mean value of e_L

$$=\frac{1}{t_1-t_0}\int_{t_0}^{t_1} e_L\times dt$$

Δe = change in percentage error e from reference value e'
$[\Delta e]_{t_0}^{t_1}$ = mean value of Δe

$$=\frac{1}{t_1-t_0}\int_{t_0}^{t_1} \Delta e\times dt$$

Δe_H = change in percentage error e_H from reference value e_H'
$[\Delta e_H]_{t_0}^{t_1}$ = mean value of Δe_H

$$=\frac{1}{t_1-t_0}\int_{t_0}^{t_1} \Delta e_H\times dt$$

Δe_L = change in percentage error e_L from reference value e_L'
$[\Delta e_L]_{t_0}^{t_1}$ = mean value of Δe_L

$$=\frac{1}{t_1-t_0}\int_{t_0}^{t_1} \Delta e_L\times dt$$

k = negative constant relating revenue loss to error $= 1/E_m t_1$
l = average loss of revenue during t_1 years due to E_m, dollars per customer
$m\alpha$, $m\beta$, $m\theta$ = mean values of α, β, and θ respectively
n = number of customers in group
P = load, watts
P_H = load at high-load test point, watts
P_i, P_j = boundaries of load subrange, watts
P_L = load at low-load test point, watts
$P_{\max}$ = maximum load taken by customer
$P_{\min}$ = minimum operating load of watt-hour meter
t = time measured from in-service date of meter, years
t_0, t', t_1 = particular values of t
y = per cent of total energy per incremental watt at load P during period t_0 to t_1 years
y_{ij} = mean value of y between ordinates P_i and P_j
y_m = mean value of y between ordinates P and $(P+\Delta P)$
ϵ = average percentage error in measurement of energy being taken by group of customers at time t
ϵ_1 = average percentage error in measurement of energy being taken by group of customers at time t_1
α, β, θ = variables
α_i, β_i, θ_i = particular values of α, β, and θ
$\Delta\beta_i$, $\Delta\theta_i$ = deviations of β_i and θ_i from $m\beta$ and $m\theta$ respectively
$\psi(P)$, $f(P)$, $f'(P)$, $f_H(t)$, $f_L(t)$, $f(t)$ = functions of variables

I. Error in Measuring Total Energy Taken Over Period of Time

The first object is to determine the error in the measurement of total energy taken by a customer over a period of one or more years of service, using the following test data:

1. The error curve of the meter determined at one time t', namely

$$e'=f'(P)$$

2. The curve of error versus time at the low-load test point of the meter during the period t_0 to t_1 years, namely

$$e_L=f_L(t)$$

3. The curve of error versus time at the high-load test point of the meter during the period t_0 to t_1 years, namely

$$e_H=f_H(t)$$

4. The consumption-versus-load characteristics of the customer during the period t_0 to t_1, expressed as

$$y=\psi(P)$$

where y = per cent of total energy per incremental watt.

The foregoing is clarified by an example. If the mean value of y between loads P and $(P+\Delta P)$ is y_m, then $y_m\times\Delta P$ per cent of the total energy is taken between loads P and $(P+\Delta P)$.

The second object is to determine the error E_m in the measurement of total energy taken by a group of customers over a period of t_1-t_0 years of service, using the data in items 1 to 4, for each customer in the group and for each customer's meter.

It is assumed that the only correlation between the meter error e and the customer's energy consumption per incremental watt y is the direct correlation due to load P. In addition, it is assumed that the meters are assigned to customers in a random fashion such that there is no correlation between the shapes of the error curves of meters and the shapes of the energy consumption curves of customers.

The error curve at any time t can be determined readily if the change in error Δe from the datum value, $e'=f'(P)$, is determined. An expression is developed for Δe.

As an approximation based on the theoretical performance of meters, it is assumed that the percentage registration can be changed only by two factors, one of which is independent of the load P, and the other varies inversely proportional to P. Then when the percentage registration is not far from 100, any change in percentage registration Δe approximately is the sum of two components, of which one is constant and the other varies inversely proportional to the load P. However, the assumptions do not apply at loads below the minimum operating load of the meter. Below the minimum operating load, $e=-100$. Thus

$$\Delta e=A+\frac{B}{P} \text{ above the minimum operating load}$$

$$\Delta e_L=A+\frac{B}{P_L}$$

$$\Delta e_H=A+\frac{B}{P_H}$$

Thus

$$A=\frac{P_H\times\Delta e_H-P_L\times\Delta e_L}{P_H-P_L}$$

$$B=\frac{P_H\times P_L}{P_H-P_L}(\Delta e_L-\Delta e_H)$$

$$\Delta e=\frac{P_H\times\Delta e_H-P_L\times\Delta e_L}{P_H-P_L}+\frac{P_HP_L(\Delta e_L-\Delta e_H)}{P(P_H-P_L)} \qquad (1)$$

For any one value of P, the values of Δe, Δe_L, and Δe_H vary with time t. Integrate each varying change in error with respect to time, and divide by the total time, t_1-t_0. Then

$$[\Delta_e]_{t_0}^{t_1}=\frac{P_H[\Delta e_H]_{t_0}^{t_1}-P_L[\Delta e_L]_{t_0}^{t_1}}{P_H-P_L}+\frac{P_HP_L\{[\Delta e_L]_{t_0}^{t_1}-[\Delta e_H]_{t_0}^{t_1}\}}{P(P_H-P_L)}$$

By definition

$$e=\Delta e+e'$$

thus

$$[e]_{t_0}^{t_1}=[\Delta e]_{t_0}^{t_1}+e'$$

$$[e]_{t_0}^{t_1}=\frac{P_H[\Delta e_H]_{t_0}^{t_1}-P_L[\Delta e_L]_{t_0}^{t_1}}{P_H-P_L}+\frac{P_HP_L\{\Delta e_L]_{t_0}^{t_1}-[\Delta e_H]_{t_0}^{t_1}\}}{P(P_H-P_L)}+e' \quad (2)$$

where

$$[\Delta e_H]_{t_0}^{t_1}=[e_H]_{t_0}^{t_1}-e_H' \quad (3)$$

$$[\Delta e_L]_{t_0}^{t_1}=[e_L]_{t_0}^{t_1}-e_L' \quad (4)$$

The total energy taken by the customer equals the area under the curve of customer's load characteristics:

$$\int y\times dP=100\%$$

The percentage error in measurement of the total energy taken for t_1-t_0 years is

$$E=\frac{1}{100}\int y[e]_{t_0}^{t_1}\times dP \quad (5)$$

This is the required solution. However, evaluation of E can be simplified.

Substitute equation 2 in equation 5. Then, above the minimum operating load

$$E=\frac{P_H[\Delta e_H]_{t_0}^{t_1}-P_L[\Delta e_L]_{t_0}^{t_1}}{P_H-P_L}+\frac{P_HP_L\{[\Delta e_L]_{t_0}^{t_1}-[\Delta e_H]_{t_0}^{t_1}\}}{100(P_H-P_L)}\times\int_{P\min}^{P\max}\frac{y}{P}\times dP+\frac{1}{100}\int_{P\min}^{P\max}e'y\times dP$$

Let

$$\frac{1}{100}\int_{P\min}^{P\max}\frac{y}{P}\times dP=a \quad (6)$$

Substitute and rearrange. Then, above the minimum operating load

$$E=\frac{P_H(1-aP_L)}{P_H-P_L}\times[\Delta e_H]_{t_0}^{t_1}+\frac{P_L(aP_H-1)}{P_H-P_L}\times[\Delta e_L]_{t_0}^{t_1}+\frac{1}{100}\int_{P\min}^{P\max}e'y\times dP$$

Below the minimum operating load

$$E=-\int_{P=0}^{P\min}y\times dP$$

Let

$$\frac{P_H(1-aP_L)}{P_H-P_L}=c \quad (7)$$

If the minimum operating load does not change with time, then over the entire load range

$$E=c[\Delta e_H]_{t_0}^{t_1}+(1-c)[\Delta e_L]_{t_0}^{t_1}+\frac{1}{100}\int_{P=0}^{P\max}e'y\times dP \quad (8)$$

To evaluate E numerically, proceed as follows:

1. Using a step-by-step method of integration, evaluate

$$\frac{1}{100}\int_{P=0}^{P\max}e'y\times dP$$

2. If a histogram is given, y is constant for each subrange of load. Then from equation 6

$$a=\frac{1}{100}\left(y_{12}\log\frac{P_2}{P_1}+y_{23}\log\frac{P_3}{P_2}+\ldots\right)=\frac{1}{100}\sum_{P\min}^{P\max}y_{ij}\log\frac{P_j}{P_i}$$

where y_{ij} is the value of y for the load subrange between P_i and P_j.

3. Evaluate c from equation 7.

4. Evaluate $[\Delta e_H]_{t_0}^{t_1}$ and $[\Delta e_L]_{t_0}^{t_1}$ from equations 3 and 4.

5. Evaluate E from equation 8.

To determine the error E_m in the measurement of the total energy taken by a group of customers, find the mean of n values of E, one value for each combination of meter and customer.

$$E_m=\frac{\Sigma E}{n}$$

where n is the number of customers in the group. From equation 8, the error pertaining to the ith customer and meter is

$$E_i=c_i[\Delta e_{H_i}]_{t_0}^{t_1}+(1-c_i)[\Delta e_{L_i}]_{t_0}^{t_1}+\frac{1}{100}\int_{P=0}^{P\max}e_i'y_i\times dP \quad (9)$$

Consider the general expression:

$$\alpha_i=\beta_i\theta_i$$

Let the mean value of

$$\alpha_i=m\alpha=\frac{\Sigma\alpha_i}{n}$$

$$\beta_i=m\beta=\frac{\Sigma\beta_i}{n}$$

$$\theta_i=m\theta=\frac{\Sigma\theta_i}{n}$$

Let $\Delta\beta_i$ and $\Delta\theta_i$ be deviations of β_i and θ_i from their mean values respectively.

$$\alpha_i=(m\beta+\Delta\beta_i)(m\theta+\Delta\theta_i)$$

$$\frac{\Sigma\alpha_i}{n}=m\beta\times m\theta+\frac{\Sigma(\Delta\beta_i\times m\theta)}{n}+\frac{\Sigma(\Delta\theta_i m\beta)}{n}+\frac{\Sigma(\Delta\beta_i\Delta\theta_i)}{n}$$

$$\Sigma\Delta\beta_i=\Sigma\Delta\theta_i=0$$

$$m\alpha=m\beta\times m\theta+\frac{\Sigma(\Delta\beta_i\Delta\theta_i)}{n}$$

If there is no correlation between β_i and θ_i, then

$$\frac{\Sigma(\Delta\beta_i\Delta\theta_i)}{n}=0$$

Thus

$$m\alpha=m\beta\times m\theta$$

To apply this result to equation 9, assume there is no correlation between:

c_i and $[\Delta e_{H_i}]_{t_0}^{t_1}$

c_i and $[\Delta e_{L_i}]_{t_0}^{t_1}$

e_i' and y_i

Then

$$E_m=c[\Delta e_H]_{t_0}^{t_1}+(1-c)[\Delta e_L]_{t_0}^{t_1}+\frac{1}{100}\int_{P=0}^{P\max}e'y\times dP \quad (10)$$

where the definitions of terms e', e_H, e_L, Δe_H, Δe_L, and y are different from those in equation 8, as follows:

e' = mean percentage error of the group of meters at load P and at time t'

e_H = mean percentage error of the group of meters at load P_H and time t

e_L = mean percentage error of the group of meters at load P_L and time t

Δe_H = change in mean percentage error e_H from reference value e_H'

Δe_L = change in mean percentage error e_L from reference value e_L'

y = mean percentage of total energy per incremental watt for the group of customers at load P, during the period t_0 to t_1 years

For the supporting equations 2 through 7, the modified definitions are:

e = average percentage error of the group of meters at load P and time t

Δe = change in the average percentage error e from reference value e'

e_H' = average percentage error of the group of meters at load P_H and time t'

e_L' = average percentage error of the group of meters at load P_L and time t'

II. Optimum Economical Service Period

Assume that the average loss of revenue, l dollars per customer, is proportional to the error E_m and the number of years of service t_1. In this instance, $t_0=0$, i.e.,

$$l=kE_mt_1$$

In a period of t_1 years, the mean annual loss of revenue is

$\frac{l}{t_1}=kE_m$ dollars per customer per year

If the meters are serviced every t_1 years at a cost of C dollars per meter per operation, then

average cost $=\frac{C}{t_1}$ dollars per year

The optimum economical service period in years occurs when

$$kE_m+\frac{C}{t_1}=\text{a minimum} \quad (11)$$

E_m is the mean error over a period of

Table III. Evaluation of $\int_{P=20}^{P=8,000} \frac{y}{P} \times dP$

Subrange P_i to P_j, Watts*	Percentage of Energy in Subrange	Percentage of Energy Per Watt y_{ij}	$y_{ij} \log (P_j/P_i)$
20 to 50	0.06	0.002	0.0018
50 to 100	1.8	0.036	0.0249
100 to 200	2	0.020	0.0138
200 to 500	10	0.0333	0.0305
500 to 1,000	26	0.052	0.0360
1,000 to 2,000	35	0.035	0.0242
2,000 to 5,000	21	0.007	0.0064
5,000 to 8,000	4	0.0013	0.0006

$$\sum_{P=20}^{P=8,000} y_{ij} \log (P_j/P_i) = 0.1382$$

* See footnote in section III of the Appendix.

time. This suggests that at any time t there exists a value of average percentage error in energy measurement ϵ, $\epsilon = f(t)$

and

$$E_m = \frac{1}{t_1}\int_0^{t_1} \epsilon \times dt$$

The optimum economical service period occurs when

$$\frac{k}{t_1}\int_0^{t_1} \epsilon \times dt + \frac{C}{t_1} = \text{a minimum}$$

This occurs when

$$\frac{d\left(\frac{k}{t_1}\int_0^{t_1} \epsilon \times dt + \frac{C}{t_1}\right)}{dt_1} = 0$$

Because

$$\frac{d\left(\int_0^{t_1} \epsilon \times dt\right)}{dt_1} = \epsilon_1$$

then

$$-\frac{k}{t_1} \times E_m + \frac{k}{t_1} \times \epsilon_1 - \frac{C}{t_1^2} = 0$$

$$k(\epsilon_1 - E_m) = \frac{C}{t_1} \tag{12}$$

Equation 12 reveals that the optimum economical service period occurs at the time t_1, when the rate of loss of revenue ($k\epsilon_1$) less the mean loss of revenue up to this time (kE_m) reaches the mean annual cost of servicing up to this time.

By definition, and from equation 10,

$$\epsilon_1 = c(e_{H_1} - e_H') + (1-c)(e_{L_1} - e_L') + \frac{1}{100}\int_{P=0}^{P\max} e'y \times dP \tag{13}$$

Using equations 10, 12, and 13, the optimum economical service period occurs when

$$kc\{(e_{H_1} - e_H') - [\Delta e_H]_0^{t_1}\} + k(1-c) \times \{(e_{L_1} - e_L') - [\Delta e_L]_0^{t_1}\} = \frac{C}{t_1}$$

$$kc\{e_{H_1} - [e_H]_0^{t_1}\} + k(1-c) \times \{e_{L_1} - [e_L]_0^{t_1}\} = \frac{C}{t_1} \tag{14}$$

With daily and seasonal changes in the errors at the high-load and low-load test points, e_{H_1} and e_{L_1} will fluctuate about a trend, and equation 14 will give a series of different values for the optimum service period t_1. It is reasonable under these circumstances to use the trend line of the measured errors to determine values of e_{H_1} and e_{L_1} for application to equation 14.

If it is desired to know whether the optimum service period has been reached at any time, a set of values for all terms may be tried in equation 14. If the resulting value of the left-hand side of the equation is less than that for the right-hand side, then it can be expected that the optimum service period will occur at a greater value of t_1.

III. Examples of Calculation of Error in Total Energy

The first example applies to one meter and one customer. The second example applies to a group of meters and customers.

Example 1. The customer's load characteristics, $y = \Psi(P)$, are those given in Fig. 3. By averaging several accuracy checks during one year, the low-load error $[e_L]_{t_0}^{t_1}$, was found to be +0.4%. Similarly, the high-load error $[e_H]_{t_0}^{t_1}$ was found to be +0.9%.

An initial measurement of the error curve of the meter indicated no error at the high-load and low-load test points. However, the meter did not operate below a load of 20 watts. Low-load tests were made at 460 watts (P_L) and high-load tests were made at 2,900 watts (P_H). Equation 8 is now applied using the steps given previously.

1. Evaluation of:

$$\frac{1}{100}\int_{P=0}^{P\max} e'y \times dP$$

Between $P=0$ and $P=20$ watts, the percentage of total energy taken by customers is −0.12%.* Above 20 watts, assume that the step-by-step integration of the error curve with the histogram in Fig. 3 gives zero error. Thus

$$\frac{1}{100}\int_{P=0}^{P\max} e'y \times dP = -0.12\%$$

2. Evaluation of:

$$a = \frac{1}{100}\int_{P=0}^{P\max} \frac{y}{P} \times dP = \frac{1}{100}\sum_{P=0}^{P\max} y_{ij} \log \frac{P_j}{P_i}$$

From Table III:

$a = 0.00138$

3. From equation 7:

Table IV. Evaluation of $\frac{1}{100}\int_{P=0}^{P=8,000} e'y \times dP$

Subrange P_i to P_j, Watts	Mean Percentage Error Between Ordinates P_i and P_j e'	Percentage of Energy in Subrange	Percentage Loss of Energy in Subrange
0 to 20	−100	0.12	0.12
20 to 50	−5.05	0.06	0.003
50 to 100	−5.35	1.8	0.096
100 to 200	−4.13	2	0.083
200 to 500	−1.4	10	0.14
500 to 1,000	−0.75	26	0.195
1,000 to 2,000	−0.6	35	0.21
2,000 to 5,000	−0.25	21	0.053
5,000 to 8,000	−0.25	4	0.010

$$-\frac{1}{100}\int_{P=0}^{P=8,000} e'y \times dP = 0.91$$

$c = 0.42$

4. In this instance:

$[\Delta e_H]_{t_0}^{t_1} = +0.9\%$

$[\Delta e_L]_{t_0}^{t_1} = +0.4\%$

5. Equation 8 reduces to:

$$E = 0.42[\Delta e_H]_{t_0}^{t_1} + 0.58[\Delta e_L]_{t_0}^{t_1} - 0.12 \tag{15}$$

Substituting the values for $[\Delta e_H]_{t_0}^{t_1}$ and $[\Delta e_L]_{t_0}^{t_1}$ gives:

$E = +0.5\%$

This is the error in measurement of the total energy taken by the customer during the year.

Example 2. The customers' average load characteristics, $y = \Psi(P)$, are those given in Fig. 3. For sample B meters, the average errors of the sample at the low-load and high-load test points were measured *in situ* periodically for 52 months. The mean value at each test point was determined for the entire period, $t_0 = 0$ to $t_1 = 52/12$ years.

At low load, the resulting mean error $[e_L]_{t_0}^{t_1}$ was −0.92%. At high load, the resulting mean error $[e_H]_{t_0}^{t_1}$ was −0.02%. The average error curve determined at one time for sample B is given in Fig. 2. From this curve

$e_L' = -0.80 \qquad P_L = 460$ watts
$e_H' = -0.27 \qquad P_H = 2,900$ watts

Thus

$[\Delta e_L]_{t_0}^{t_1} = -0.12\%$

$[\Delta e_H]_{t_0}^{t_1} = +0.25\%$

As in example 1, $c = 0.42$.

During the 52-month period, assume that the minimum starting load of the meters in the sample was 20 watts. Then the error curve in Fig. 2, $e' = f'(P)$, must be integrated with the customers' consumption characteristics in Fig. 3, $y = \Psi(P)$, between zero and 8,000 watts. This is done step by step in Table IV. The result is

$$\int_{P=0}^{P=8,000} e'y \times dP = -0.91\%$$

This includes the percentage error due to the energy taken between 0 and 20 watts = −0.12%.

---

* In Fig. 3, the first load subrange, 0 to 100 watts (which includes 2% of the total energy), has arbitrarily been divided again into three narrower load bands, 0 to 20, 20 to 50, and 50 to 100 watts respectively. The amounts of energy taken in these respective bands are as 0.04 to 0.02 to 0.6. These ratios are based on the data for very low loads shown in Fig. 5. In this figure, it is assumed that 0.6% of the total energy is taken between 50 and 100 watts.

Equation 10 reduces to

$$E_m = 0.42[\Delta e_H]_{t_0}^{t_1} + 0.58[\Delta e_L]_{t_0}^{t_1} - 0.91$$

Substituting the values for $[\Delta e_H]_{t_0}^{t_1}$ and $[\Delta e_L]_{t_0}^{t_1}$ gives:

$$E_m = -0.87\%$$

This indicates that the total energy taken by customers having meters similar to those in sample B was registered low by approximately 0.9% in a period of 52 months.

IV. Samples of Calculation of Optimum Service Period

Because groups of meters tend to have increasing minus errors as the length of service increases, the first example applies to a uniformly increasing error for the group. The second example applies to sample B meters.

Example 1. With uniformly increasing error:

$$\epsilon = bt$$

From equation 11, the optimum economical service period is t_1, where

$$k\left(bt_1 - \frac{bt_1}{2}\right) = \frac{C}{t_1}$$

$$t_1 = \sqrt{\frac{2C}{kb}} \text{ years} \qquad (16)$$

If the increase in error ϵ/t is −0.1% per year, then $b = -0.1$. If the average annual revenue per customer is $80, and if the percentage loss of revenue per year is seven tenths of the minus percentage error, then $k = -0.56$. If the cost of servicing C is $6.00 per operation per meter, then from equation 16 the optimum service period t_1 is 14.6 years.

Example 2. Equation 14 is applied using the measured data obtained on sample B meters, and using the histogram shown in Fig. 3. A test will be made to find whether the optimum economical service period has been reached after 52 months of service.

Assume that the dashed lines in Fig. 1(B) represent the trend curves to be used to determine values for e_{H_1} and e_{L_1}, at time $t_1 = 52/12$ years. From Fig. 1(B):

$e_{H_1} = -0.17\%$ $\qquad P_H = 2{,}900$ watts

$e_{L_1} = -1.20\%$ $\qquad P_L = 460$ watts

As in example 2 of section III of the Appendix:

$$[e_H]_{t_0}^{t_1} = -0.02\%$$

$$[e_L]_{t_0}^{t_1} = -0.92\%$$

$$c = 0.42$$

Assume that $k = -0.56$ and $C = \$6.00$, as in the foregoing example. Substitute values in equation 14.

left-hand side = $0.126 per customer per year

right-hand side, $C/t_1 = \$1.38$ per meter per year

Because the left-hand side is less than C/t_1, the optimum economical period has not been reached.

From equation 12, the left-hand side of equation 14 equals $k(\epsilon_1 - E_m)$. Suppose the errors of sample B meters continue to increase such that $k(\epsilon_1 - E_m)$ remains constant and equal to 0.126, then the optimum service period will be

$$\frac{C}{k(\epsilon_1 - E_m)} = 48 \text{ years approximately}$$

References

1. Central Electric Stations. Dominion Bureau of Statistics, Ottawa, Ont., Canada, 1955.

2. Statistics of Electric Utilities in the United States, 1955; Privately Owned; Publicly Owned. *Publications S-124, S-125,* Federal Power Commission, Washington, D. C.

3. Meter Problems and Consumers' Load Characteristics, L. B. S. Golds, P. Schiller. *Proceedings,* Institution of Electrical Engineers, London, England, vol. 101, pt. II, Dec. 1953, pp. 619–31.

4. Code for Electricity Meters. *C12-1941,* American Standards Association, New York, N. Y., fourth edition, p. 93.

5. Industrial Experimentation (book), K. A. Brownlee. Chemical Publishing Company, Brooklyn, N. Y., 1952, pp. 32–26.

6. New Sampling Technique Cuts Meter Testing Cost, W. C. Harrison. *Electrical World,* New York, N. Y., July 15, 1957, pp. 72–74.

7. Modern Elementary Statistics (book), J. E. Freund. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1952, sects. 6.2, 6.3.

8. Biometrika Tables for Statisticians, Vol. I (book), E. S. Pearson, H. O. Hartley. Cambridge University Press, Cambridge, England, 1956, Table 41.

9. Two Numerical Indexes Evaluate Masses of Meter Test Data, F. H. Busch. *Electrical World,* Dec. 18, 1950, pp. 84–86.

10. *Advance Reports,* Engineering Section, 60th Annual Convention, Canadian Electrical Association, Montreal, Que., Canada, p. 48.

◆

Discussion

L. B. S. Golds (Eastern Electricity Board, Ipswich, England): Economics of metering is a complicated subject and I agree that it is helpful and indeed necessary to employ statistical techniques to bring the mass of data into a manageable form. One must, however, be careful that such techniques are not pressed to fulfill a function which is beyond their capabilities.

Mr. Vanderleck suggests that the service period of meters should be determined by the percentage of meters registering outside the statutory limits. This is one of the criteria upon which the period might be based, but up to the present it does not appear that a statutory authority would recognize officially that any meter may register energy outside the legal tolerances; although, from a practical and theoretical standpoint, it must of course be recognized that such a possibility exists. The difficulty lies in making statutes to fit statistical theory.

Referring to the assessment of the percentage of meters registering the energy outside the tolerances it is stated in the paper that this percentage may be obtained by a count of the meters in the sample, and when compared with the calculated figure derived from the knowledge of the average error and standard deviation there is an

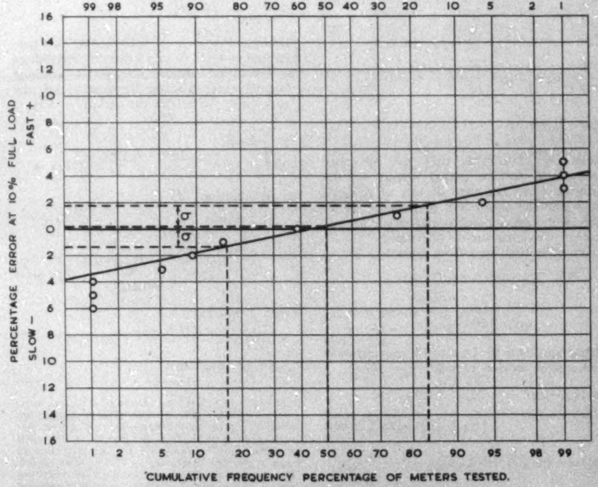


Fig. 7. Cumulative percentages of off-circuit errors

appreciable difference. The author, I believe, suggests that the count is the "true" fraction. Is this "true" figure really true, or is not the true figure obtained only by testing the whole population?

Further, the author arrives at the conclusion that the frequencies of errors of meters tested after return from service are not distributed normally. This conclusion will, of course, apply to a sample drawn from a mixed "population," but I cannot agree with the conclusion when the sample is drawn at random from the same population. In Fig. 7 the cumulative percentages of off-circuit errors of a certain type and make of meter are plotted on probability graph paper. These meters had been in service for approximately 14 years and it may be seen that normality is indicated between 5% and 95% of the meters tested. As would be expected the frequency with which errors occur exceeding 5% and 95% of the total will be so low that very large numbers of meters would be needed to show that the errors were not normally distributed. This subject was dealt with by Josephs in 1943.[1]

I may have misinterpreted the figures in the paper but it appears that the high- and low-load test points are fairly close together on the curve of the meter, i.e., 2,900 watts and 460 watts, and the reason for selecting these points is not clear to me.

The author points out rightly that the error of total energy measurement is affected by the choice of rating, type, make, and age of the meter for a particular service. The proper choice can increase the service life of the meter materially and result in meters being available for a longer period of service leading to a saving of capital expenditure upon new meters.

I am naturally gratified that the investigation which I initiated in the United Kingdom has been pursued by the author and trust that the importance of efficient metering will be generally recognized in the future in other countries.

Reference

1. H. J. Josephs. *Journal*, Institution of Electrical Engineers, London, England, vol. 92, pt. 2, 1943.

J. M. Vanderleck: In reply to Mr. Golds, the literature[1] reveals that the regulative authorities in six states of the United States have authorized the use of sampling procedures in place of periodic testing of all meters. Such procedures permit a small percentage of meters to operate with relatively large errors.

With respect to the fraction of meters registering outside specified error tolerances, the paper discusses the true fractions and normal fractions of six groups of meters with 1,150 to 6,214 meters per group. In no instance was the true fraction intended to be taken as that of a population represented by the group as a sample; that is, the groups are considered as entire populations.

Possibly all six groups are mixed populations as Mr. Golds suggests. If so, then a utility's problem is to sort most of the meters into "pure" populations, these having normal distributions by definition. The curve given by Mr. Golds appears inconclusive because of the uncertain distribution of errors at the two tails of the curve. Since the percentage of meters outside of the tolerances are the essential data, the tails are important.

The values of low load and high load chosen, namely 460 and 2,900 watts respectively, were those prescribed by the regulative authority for the ratings of the meters considered in the paper.

Reference

1. Utilities in Six States Adopt Sample Testing of Meters. *Electrical World*, New York, N. Y., Mar. 10, 1958, pp. 46–47.

Thermal Impedance of Cooling Fins

E. J. DIEBOLD
MEMBER AIEE

WERNER LUFT
NONMEMBER AIEE

Synopsis: Cooling fins of equal thickness and known heat transfer have a definite thermal impedance which is hard to determine analytically. Computing generally valid curves of normalized variables eliminates most of this analysis and establishes simple relations for design parameters and impedance. Optimum design proportions appear in the analysis, eliminating the trial-and-error approach to find them. Comparative analysis of cooling systems is facilitated.

SEMICONDUCTOR DEVICES have widespread acceptance because of their ability to handle large amounts of power although they are extremely small. Having little cooling surface of their own, effective heat dissipation is possible only by conduction to a heat exchanger whenever an appreciable amount of heat is generated. Maximum working temperature of a device is usually determined by the limit of reliable performance, whereas the ambient temperature and the cooling medium depend on the application and may assume rather extreme conditions in military gear.

Devices without the integral heat exchanger usually have a substantial mounting stud, which is advantageous for cooling by bolting to a metal plate which acts as a cooling fin. Under given conditions of heat generation, maximum permissible device temperature, ambient temperature, air pressure, radiant heat from other sources, available air flow, allowable space, and multiple connections, the designer should find an efficient and economic cooling fin, preferably using standard materials. Considering all these limitations, the designer may have to settle for the best possible solution under the circumstances. The present analysis has the objective to facilitate this, using a minimum of numerical work and giving a best possible insight into the potential solutions.

Generally, the subject of cooling fins is well covered in the literature.[1-3] (The bibliographies of references 2 and 3 are of value.) However, because of the many different applications of cooling fins, the available literature is often too broad to be of immediate use for cooling semiconductors. The objective, then, is to narrow down the analysis to specific problems, by using the following restrictions:

1. Results of the mathematical analysis are presented as curves covering a wide range but without extreme accuracy, knowing that many design parameters are too inaccurate to permit a more careful analysis and assuming that designers are unwilling to use lengthy numerical methods to gain a fictive accuracy.

2. Results are normalized, using the base of the device to be cooled as reference dimensions.

3. Geometric restrictions consist of a thin, wide fin with equal thickness, and constant cooling conditions over the whole area.

Within these restrictions the analysis offers basic concepts of general validity permitting direct comparison of widely different versions. Depending upon the design problem (e.g., improvement of existing devices, planning of entirely new equipment, comparison of competing versions) different modes of approach are needed. The analysis therefore shows several different concepts for the solution of the same problem.

Method of Approach

Heat generated in the active part of a device flows through different paths into

---

Paper 58-926, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 12, 1958; made available for printing April 23, 1958.

E. J. Diebold and Werner Luft are with the International Rectifier Corporation, El Segundo, Calif.

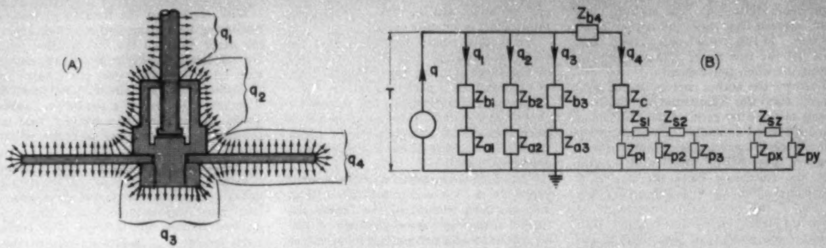


Fig. 1. A—Schematic representation of heat dissipation from a semiconductor device. B—Equivalent electric circuit of (A)

the ambient space. Comparing the flow of heat to an electric current, the temperature rise over ambient to an electric potential above ground and the thermal impedance to the electric resistance, an electric circuit model can be made. Fig. 1(A) shows an example of heat dissipation, and 1(B) the equivalent electric circuit. Heat is generated at the rate q per second; it causes a temperature rise T in the heat generator, and is dissipated along several paths to the ambient.

Heat flows away through the lead at the rate q_1; it encounters the internal thermal impedance Z_{b1}, and the external heat transfer impedance Z_{a1}. Similarly, heat flows as q_2 through the body of the device above the fin and as q_3 through body and hardware below the fins. Corresponding impedances are marked internally as Z_{b2}, Z_{b3}, and externally as Z_{a2}, Z_{a3}. If the device is tested without fin, these series-parallel connected impedances are measured as a lump figure Z_d.

If a cooling fin is attached to the device, a major part of heat flows as q_4 through the internal impedance Z_{b4} and the contact impedance Z_c into the cooling fin. These two impedances may be determined as a lump sum Z_i.

Lumped impedances of the device, as determined in the foregoing, are always approximately the same (except for a temperature correction). It is worth while to determine these impedances accurately as an average from many samples. For the cooling fin, however, the material, the kind of cooling, and the shape, size, and thickness may vary, and unless a good analytical method is available any new cooling fin or alteration on an existing one represents a testing problem. Accordingly, this analytical method concentrates on the cooling fin alone.

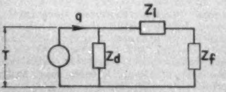


Fig. 2. Simplified equivalent circuit of heat dissipation

Fig. 1(B) illustrates the problem of the distributed series and parallel thermal impedances which is particular to cooling fins. As the heat penetrates into the fin and spreads out, it encounters an infinity of infinitely low (but different) series impedances Z_{s1}, Z_{s2}, etc.; they decrease as the radius increases. Simultaneously, heat escapes from both sides of the fin through the parallel impedances Z_{p1}, Z_{p2}, etc., which are infinitely high but which decrease relatively as the radius increases. These impedances are represented as an interrupted chain in Fig. 1(B). There is no elementary way to find a lump impedance Z_f to replace such an infinite chain of varying elements; the following analysis, however, solves the problem by the use of generally valid data in the form of graphs.

The thermal impedance problem then resolves to a simple model shown in Fig. 2, wherein the heat dissipation impedance Z_d of a device is found by operating it alone in the cooling medium to be used. To find the internal impedance Z_i, the device is fastened to a heat sink of known impedance and the over-all impedance is determined. Knowing also the values of the impedances Z_d and Z_f, Z_i can be calculated. To analyze a new problem the known impedances Z_d and Z_i are combined with a fin impedance Z_f to form a new over-all impedance.

ASSUMPTIONS

Semiconductor devices and their cooling fins operate under various ambient restrictions, which prevent accurate determination of thermal impedances, except in very special cases. Accordingly, the analysis may be made, based on rather summary assumptions, with the knowledge, however, that the results should be applied with caution. The following assumptions give good practical results, particularly in comparative analysis; their validity is discussed in reference 1.

1. Heat transferred from metallic fin to air is proportional to the difference of temperature, and the factor of proportionality is the same over the whole fin surface. Since it is, in effect, neither exactly proportional nor actually constant, an average value must be determined for the different temperatures, temperature differentials, and locations on the fin. In extreme cases (such as of boiling fluids, or cooling at high temperature in vacuum) this theory offers only rough approximations. Heat transfer coefficients (radiation, convection, forced air) are shown in Appendix I and reference 4).
2. The fin is plane, of unvarying thickness, and homogeneous.
3. The fin is of circular symmetry around the heat source.
4. The fin thickness is small compared to its width and the thermal impedance for heat flow across the fin thickness is negligible.
5. Temperature and heat flow are independent of time.
6. No heat is generated in the fin.
7. The ambient temperature is constant.
8. The inner diameter (of the heat source) is much smaller than the outer diameter of the fin.
9. Heat dissipated from the outer edge of the fin is neglected, or considered by widening the fin by half its thickness.
10. In multiple-fin arrangements the fin spacing is greater than two boundary layers of coolant flow (see discussion in reference 2).

Assumptions 2, 4, 8, and 10 are specific to the application of fins for cooling semiconductor devices, and for this reason the theory here may contradict other theories on cooling fins, although it is correct within the assumptions.

Assumption 3 is most restrictive, and rarely met. It is tolerable because deviations from this condition usually occur

Vol.
Pt. 1
1958

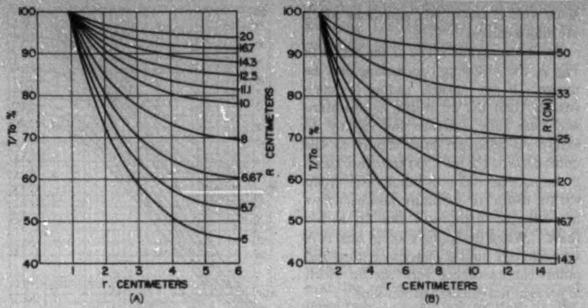


Fig. 3. Temperature–distance curves for fin with 1-cm inner radius, varying outer radius, and several natural radii

A—6.3-cm outer radius

B—15.2-cm outer radius

at an appreciable distance from the heat source. As shown in Fig. 3(A), the temperature distance variations in the outer region of the fin are small; hence it is rather unimportant whether the fin is circular, square, or rectangular. If the heat source is located on or close to the edge of the fin, it is advisable to determine the thermal impedance of a larger symmetrical fin, which contains the one under consideration, and to multiply it by the ratio of the symmetrical over the unsymmetrical area.

NOMENCLATURE

a, b = length of sides of cooling fins, m
A = heat transfer area, m^2
A_r = heat conduction cross-section area (radius r), m^2
C = degrees centigrade
d = diameter circular cooling fin, m
D_b, D_h = diameter of base and fin hole, m
Φ = fin effectiveness: ratio fin to base admittance
h = heat transfer coefficient, w/m^2C (watts per square meter, C); see Appendix I
H, K = Hankel functions, real and imaginary argument
η = fin efficiency: admittance ratio, real fin to infinitely conductive fin
J, I = Bessel function, first kind, real and imaginary argument
k = thermal conductivity of fin material, w/mC
m = meter
Λ_e = normalized volume of fin, excluding center hole
Λ_i = normalized volume of fin, including center hole
ψ = thermal admittance function of cooling fin
q = heat flow, w
q_0 = heat flow at inner radius, w
q_1, q_2, q_3, q_4 = heat flow through lead, upper body, lower body, fin, w
r = radius of any segment in fin, m
r_0, r_1 = inner and outer fin radius, equivalent for analysis, m
$R=\sqrt{ks/2h}$ = natural radius, m
R_t, R_q = subtangents to temperature and heat curves, m
s = thickness of fin, m
S_t, S_q = tangents to temperature and heat curves, m
$w=r_1-r_0$ = width of fin, m
T = temperature rise over ambient, C
T_0, T_1 = temperature rise at inner and outer radius of fin, C
θ_1, θ_2 = integration constants, C
$u=r/R$ = normalized variable
V_e = volume of fin excluding center hole, m^3
V_i = volume of fin including center hole, m^3
w = watt
$Z=T/q$ = thermal impedance, C/w
Z_{a1}, Z_{a2}, etc. = external thermal impedances, C/w
Z_{b1}, Z_{b2}, etc. = internal thermal impedances, C/w
Z_d, Z_i, Z_f = equivalent lumped thermal impedances: device, internal-to-fin, fin, C/w
Z_{s1}, Z_{s2}, etc. = infinitesimal thermal impedances in series in fin, C/w
Z_{p1}, Z_{p2}, etc. = infinite thermal impedances in parallel to fin, C/w

Fig. 4(A) and (B) shows a device which is cooled by a square fin; its equivalent circular fin is shown in (C). The inner radius of the fin is given approximately by the mean value of the radii of the base of the device and the hole in the fin.

$$r_0=\frac{D_h+D_b}{4} \tag{1}$$

Fins with a pipe-thread hole have an inner radius equal to half the pitch diameter of the thread. The diameter of the device to be cooled is assumed known, hence the inner diameter r_0 of the fin is taken as reference diameter.

The external radius of the fin is determined by the following:

For a square, sides b, thickness s

$$r_1=\sqrt{\frac{b^2+2bs}{\pi}} \tag{2}$$

For a rectangle, sides a, b

$$r_1=\sqrt{\frac{ab+(a+b)s}{\pi}} \tag{3}$$

For a circular fin, diameter d

$$r_1=\frac{d}{2}\sqrt{1+\frac{2s}{d}} \tag{4}$$

For studying the heat transfer in Fig. 4(C) and with the foregoing dimensions, only the two faces of the fin which are shown by heavy solid lines are considered. For a first approximation the thickness of the fin in these equations may be neglected.

Heat Flow and Temperature Rise

Fig. 5(A) shows a cross section of the fin and (B) the curves of temperature and heat flow, plotted against the same radius. At an arbitrary distance from the center, the radial heat flow q passes through a cylindrical area A_r into the annual volume element dV, bounded by the radial increment dr. Heat conductivity k of the fin and heat transfer coefficient h to the ambient space are constant.

The temperature drop in the volume element dV is given by the law of conduction, as

$$dT=-\frac{q\,dr}{2\pi rsk} \tag{5}$$

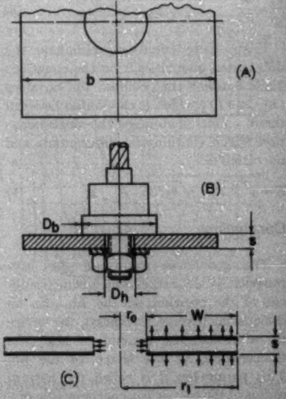


Fig. 4. Comparison of a real device and its cooling fin and the simplified fin for mathematical treatment

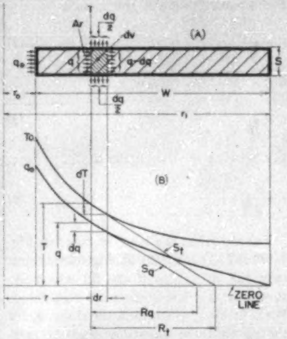


Fig. 5. Cross section of simplified fin and temperature: heat flow curves along a radius of the fin

Heat dissipated by the volume element is given by the law of heat transfer:

$$dq = -4\pi hTrdr \tag{6}$$

If in Fig. 5, the tangent S_t is drawn to the temperature curve (at the radius r), and the tangent Sq is drawn to the heat flow curve (as the same radius), the distances from the intersections of these tangents with the base line ($T=0$, $q=0$) to the radius r are defined as the subtangents R_t and Rq. By the similarity of triangles the following is obtained.

$$\frac{dT}{T} = \frac{dr}{R_t} \tag{7}$$

$$\frac{dq}{q} = \frac{dr}{R_q} \tag{8}$$

Using these equations, eliminate the differential quantities from the equations 5 and 6, which then contain the variables only as rT/q which is eliminated between them. Thus is obtained the single equation which contains no differentials and no variables:

$$R = \sqrt{R_tR_q} = \sqrt{ks/2h} \tag{9}$$

Conclusions

The geometric mean of the subtangents is a constant which is independent of the position on the fin, the fin dimensions (except thickness), the device to be cooled, the amount of heat, the temperature, or the shape of the fin. This parameter R is called the natural radius of the fin; it is a characteristic distance of a certain sheet material and heat transfer mode. For many investigations it is advantageous to normalize all dimensions with R and obtain equations which are dimensionless.

By observing the shape of the curves in Fig. 5, the following conclusions may be drawn: if the natural radius is much smaller than the fin width, temperature and heat flow decrease too rapidly to assure an adequate heat supply to the outer parts of the fin. On the other hand, if the natural radius is much larger than the fin width, temperature and heat flow decrease too little, and the fin is too thick. When designing cooling fins, it is advantageous first to determine a natural radius commensurate with the fin width to prevent unrealistic assumptions.

For constant values of thermal conductivity and heat transfer constants, the model law of cooling fins is that the thickness s varies with the square of the lineal dimensions, e.g., doubling the fin size means quadrupling the thickness if a proportional flow is expected.

Solution of the Differential Equations

Equations 5 and 6 may be combined into a single equation, containing either the temperature T or the heat flow q only. Equation 5 is rewritten and differentiated to

$$\frac{dq}{dr} = -2\pi ks\left(\frac{dT}{dr} + r\frac{d^2T}{dr^2}\right) \tag{10}$$

This derivative is also available from equation 6; it combines with 10 to form

$$\frac{d^2T}{dr^2} + \frac{1}{r}\frac{dT}{dr} - \frac{2h}{ks}T = 0 \tag{11}$$

This is a Bessel differential equation, order zero; the solution is shown in Appendix II. Temperature distance relationship is given there by equation 41 as a function of the normalized values: r/R, r_0/R and r_1/R.

In the most general case the reference radius r is the only variable but the parameters r_0, r_1 and R are unknown. Even when working with normalized variables (with respect to a yet undetermined value of R) there are still two unknown parameters. A general plot of the temperature-distance function therefore involves at least two independent parameters and one variable; this cannot be presented in one graph. For this reason only two specific examples are shown, in which the inside radius r_0 is held constant at 1 cm (centimeter) and the outside radii are 6.3 cm for Fig. 3(A) and 15.2 cm for (B). Both figures are plotted for several values of the natural radius R. Temperature-radius curves follow a common pattern: the temperature drop within the fin is highest near the inner radius and approaches zero at the outer radius. In Fig. 3(B), if the natural radius is increased from 16.7 cm to 50 cm, thus making the fin nine times heavier, the outer rim temperature increases from 50% to 90% of the inner rim temperature. In other words, 40% of temperature range is gained by the addition of 800% more material. To find an optimum of heat dissipation versus fin volume other concepts must be introduced which facilitate the comparison.

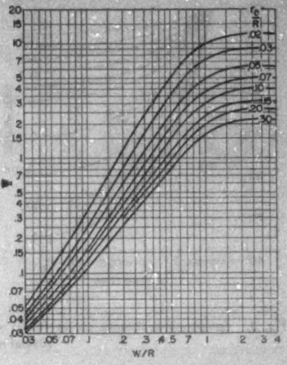


Fig. 6. Thermal admittance function of cooling fins plotted against ratio of fin width to natural radius and several ratios of inner to natural radii

Total heat dissipating capacity of the fin, for a certain temperature rise at the point of heat influx, may be found from equation 43 as

$$q_0 = 4\pi hr_0RT_0\psi \tag{12}$$

with

$$\psi = \frac{I_1\left(\frac{r_1}{R}\right)K_1\left(\frac{r_0}{R}\right) - I_1\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)}{I_0\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right) - I_1\left(\frac{r_1}{R}\right)K_0\left(\frac{r_0}{R}\right)} \tag{13}$$

Define ψ as the thermal admittance function of the cooling fin; it indicates the ability of the fin to dissipate heat for a given inner radius, natural radius and temperature rise. All further discussion will be based on this function, which eliminates the necessity of using an unwieldy expression. Fig. 6 shows numerical values of the thermal admittance function for a wide range of values, expressed as fin width over natural radius and inner radius over natural radius. In this figure it is apparent that the ratio of fin width over natural radius is important; if it is larger than unity the thermal ad-

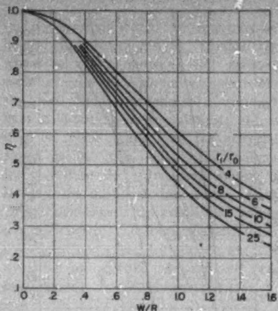


Fig. 7. Fin efficiency plotted against ratio of fin width to natural radius and several ratios of inner to natural radii

mittance is almost constant, no matter how wide the fin is made.

From equation 12 the thermal impedance of the fin is determined as

$$Z_f = \frac{T_0}{q_0} = \frac{1}{4\pi h r_0 R \psi} \tag{14}$$

With this equation and Fig. 6, the thermal impedance of any fin may be determined without much numerical work.

Fin Efficiency

The fin efficiency may be defined as the ratio of total heat dissipated by a cooling fin to heat dissipated by an identical fin of infinitely conductive material. It is a measure for the loss of heat dissipation capacity by the internal temperature drop of the fin.

Equation 12 gives the heat dissipated by the real fin. Heat dissipated by a fin of infinite conductivity may be expressed as

$$q_t = 2\pi(r_1^2 - r_0^2)hT_0 \tag{15}$$

$$\eta = \frac{q_0}{q_t} = \frac{2(R/r_0)\psi}{(r_1/r_0)^2 - 1} \tag{16}$$

Fig. 7 shows the efficiency plotted against the ratio of fin width over natural radius, for several values of outer to inner radii. The efficiency is greatly influenced by the relative fin width and, as in Fig. 6, it appears that relative widths of more than unity are not economical. Very low relative widths produce efficiencies close to unity, very desirable, of course, but also wasteful in material. Fig. 8 shows a plot of efficiency versus the ratios of natural over inner, and of outer over inner, radii. It covers a wide range with better accuracy than Fig. 7, though it is less informative at first sight. Knowing the efficiency one may calculate the heat dissipated by the fin as

$$q_0 = hA\eta T_0 \tag{17}$$

For noncircular fins, the efficiency may be approximated by selecting a safe (i.e., large) outer radius, then the real heat dissipation will be somewhat better than calculated.

In an investigation where any dimension, including the inner radius, may be changed, the concept of fin efficiency is most valuable, although it does not present an optimum value. When, however, the optimum curves (for constant inner radius of the fin) derived in Fig. 9(A) and (B) are compared with the curves for constant efficiency in Fig. 8, it is found that they follow very closely the constant efficiency curve of 0.5. At first sight this result is surprising because the mathematical derivation of the constant efficiency curve is not too difficult (from equation 16), whereas the derivation of the optimum curve defies direct analysis and must be made graphically. However, if this result is analyzed by the practical engineer, it is obvious, and it may be concluded that optimum fins have an efficiency of approximately 50% which corresponds to equal internal and external thermal impedances.

Fin Effectiveness

For fins with constant inner radius the concept of fin effectiveness is more useful than the fin efficiency. The ratio of heat dissipated by the cooling fin to the heat dissipated by the base area (inner radius) under the same conditions and for the same temperature rise is defined as fin effectiveness.

Again the reference value is fictive: It is part of an inseparable total, and the heat transfer constant (of the device alone) is different from the constant for the cooling fin. However, a close approximation may be found by assum-

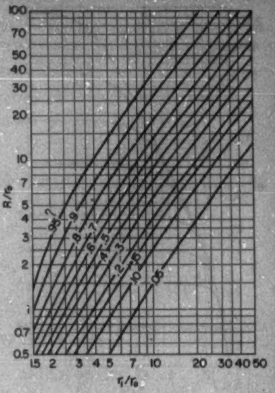


Fig. 8. Curves of constant fin efficiency plotted against ratios of outer over inner, and natural over inner, radii

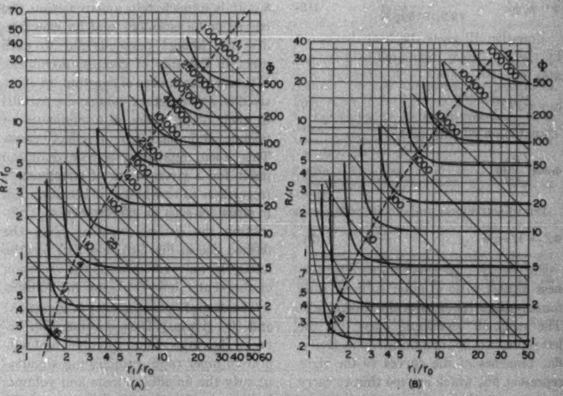


Fig. 9. Curves of constant fin effectiveness and constant per-unit volume plotted against ratios of outer over inner, and natural over inner, radii

A—Including center hole
B—Excluding center hole

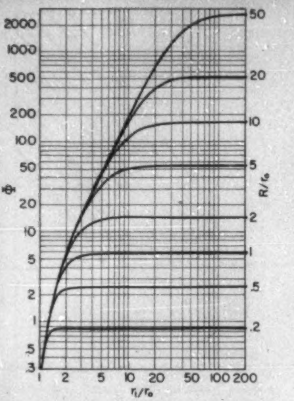


Fig. 10. Fin effectiveness plotted against ratio of outer to inner radii for several values of natural to inner radii

ing that the heat dissipated per unit area of the device is constant and by multiplying it by the computed ratio of heat transfer constants obtained from known information; see Appendix I.

If the heat is known which can be dissipated by the device alone, the heat q_b dissipated by the base area alone (equal heat transfer constant as the fin) may be determined. The heat q_0 to be dissipated by the fin is also known.

Fin effectiveness is

$$\Phi=q_0/q_b \tag{18}$$

Equation 12 gives the heat dissipated by the fin. Heat dissipated by the base area is

$$q_b=\pi r_0^2 h T_0 \tag{19}$$

Thus the fin effectiveness is

$$\Phi=4\left(\frac{R}{r_0}\right)\psi \tag{20}$$

Fin effectiveness and fin impedance is

$$\Phi Z_f=1/\pi r_0^2 h \tag{21}$$

Fig. 10 gives the curves of fin effectiveness against ratio of outer to inner radii for several ratios of natural to inner radii. The common ascendent curve on the left represents fins of infinite conductivity, flat branches of the curves to the right represent fins which are too thin to carry heat towards their outer areas.

In designing a new cooling fin, the thermal impedance Z_f which must be achieved is usually known. From equation 21 the desired effectiveness is obtained and Fig. 10 yields a certain number of possible ratios of dimensions, the ratio of natural over inner radius being almost independent of the ratio of outer to inner radius.

Another plot for the values of fin effectiveness is shown in Fig. 9(A) and (B). In these curves, the value of effectiveness is kept constant, i.e., each curve represents many fins which can perform the same cooling duty. Similarly, the figure contains also the curves of constant volume to permit direct comparison of fin effectiveness and volume. The fin volume is defined as V_t or V_e, either including or excluding the material punched out for the center hole.

$$V_t=\pi r_1^2 s=2\pi\frac{h}{k}r_0^4\left(\frac{r_1}{r_0}\right)^2\left(\frac{R}{r_0}\right)^2 \tag{22}$$

$$\begin{aligned}V_e&=\pi(r_1^2-r_0^2)s\\&=2\pi\frac{h}{k}r_0^4\left[\left(\frac{r_1}{r_0}\right)^2-1\right]\left(\frac{R}{r_0}\right)^2\end{aligned} \tag{23}$$

Per-unit fin volume (Δ_t including, Δ_e excluding center hole) is defined as

$$\Delta_t=\frac{V_t k}{2\pi h r_0^4}=\left(\frac{r_1}{r_0}\right)^2\left(\frac{R}{r_0}\right)^2 \tag{24}$$

$$\Delta_e=\frac{V_e k}{2\pi h r_0^4}=\left[\left(\frac{r_1}{r_0}\right)^2-1\right]\left(\frac{R}{r_0}\right)^2 \tag{25}$$

These values are only representative as long as the inner radius is constant; if it varies, the volume is proportional to the ratio Δ times the fourth power of the inner radius. In Fig. 9(A), fin effectiveness and per-unit volume are plotted against the same ratios of radii as in Fig. 8. It is immediately apparent that the curves of fin efficiency and fin effectiveness are almost perpendicular to each other in the area of greatest usefulness. Wherever the curves of constant fin effectiveness and constant volume are parallel to each other, there is a minimum of volume and maximum of effectiveness. This, then, represents a true optimum point for a given value of effectiveness (regardless as to whether it is economical to strive for such an effectiveness). Connecting the optima together gives the dashed line in Fig. 9(A) and (B); in the latter the only difference is the exclusion of the center hole from the fin volume.

The dashed optimum lines in Figs. 9(A) and (B) give a preferred connection between the ratios of outer to inner and natural to inner radii, establishing simultaneously the fin effectiveness and volume. It is interesting that the influence of the center hole is very small, the optimum curves in (A) and (B) being almost identical; (A) is more interesting from the standpoint of material cost, and (B) from the standpoint of weight. Curves in these two figures are deleted in the upper left-hand corner because the basic assumption of thin fins is invalid.

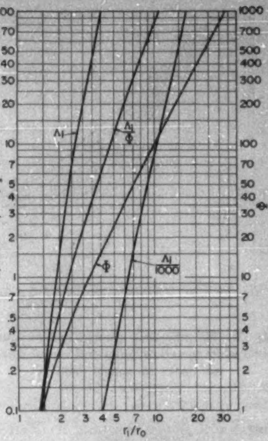


Fig. 11. Per-unit fin volume, fin effectiveness and ratio of volume over effectiveness along optimum curve of Fig. 9(A), plotted against ratio of outer to inner radii

Fig. 11 shows the curves of normalized volume and effectiveness plotted against the ratio of outer to inner radii, along the optimum curve of Fig. 9(A); hence these values are correct only for the ratios of dimensions established by the dashed line in 9(A). Forming the ratio of fin volume over effectiveness gives

$$\frac{\Delta_t}{\Phi}=\frac{k}{2r_0^2}V_t Z_f=\frac{1}{4\psi}\left(\frac{r_1}{r_0}\right)^2\left(\frac{R}{r_0}\right) \tag{26}$$

This ratio may be defined as a "figure of demerit;" it indicates how much material is needed to achieve a certain effectiveness (the inner radius and conductivity unchanged). Surprisingly, this figure of demerit increases rapidly with increasing ratio of outer to inner radii and hence increasing effectiveness.

Conclusions

Fins are much more wasteful in material if they are wide; in the range of practical usefulness the volume increases approximately with the third power of the ratio of radii and the power 4/3 of the effectiveness. Also, many narrow fins are much lighter than one wide fin.

Unfortunately, this conclusion conflicts with the realities of present design and applications, but this insight may be helpful in future designs, particularly where saving of material and space is essential. In multiple-fin arrangements the mass of material required to conduct heat from the device to a plurality of fins may be much higher than the material saved in the fins. Economic analysis incorporating this additional volume is beyond the scope of this paper; it is, easily performed by using the values available herein, in conjunction with other considerations given by the device, the available space, and the preferred proportions.

Appendix I. Heat Conductivity and Heat Transmission

Heat conductivity k for some metals in w/mC (approximately) is as follows:

Copper, 370
Aluminum, 200
Brass, 100
Mild steel, 45

Heat transmission may be analyzed as follows.

Radiation:

h_r in (w/m^2C), ϵ = emissivity

$$h_r=5.71\,10^{-8}\epsilon\left(\frac{T_1^4-T_2^4}{T_1-T_2}\right)$$

$$T_a=\frac{T_1+T_2}{2};\quad T_g^2=\frac{T_1^2+T_2^2}{2};\quad T_g\simeq T_a$$

$$h_r=2.28\,10^{-7}\epsilon T_aT_g^2\simeq 2.28\,10^{-7}\epsilon T_a^3$$

Note: T_1, T_2, T_a, T_g are expressed in degrees Kelvin

The foregoing is valid only if radiation is not obstructed by other hot bodies (cooling fins).

Free convection (in air at sea level):

h_c in w/m^2C, approximately

$$h_c=1.4\left(\frac{\Delta T}{L}\right)^{1/4}$$

where

ΔT = temperature difference metal-air, C
L = height of fin in meters (or length of flow along fin)

Forced cooling (in air):

h_f in w/m^2C

$$h_f=0.055\frac{k_a}{L^{1/4}}\left(\frac{V}{\nu}\right)^{3/4}$$

T (C)	100 k_a	$10^5\,\nu$
20	2.6	1.5
60	2.9	1.9
100	3.2	2.4
140	3.5	2.8

where

k_a = thermal conductivity of air, w/mC
ν = cinematic viscosity, m^2/s
V = air velocity, m/s

Minimum fin spacing is approximately two thicknesses of boundary layer b, or slightly less:

$$b=5.83\sqrt{\frac{\nu L}{V}}$$

Appendix II. Solutions of Equations

Introduce the new variable $u=r/R$, and insert it into equation 11:

$$u=r/R \tag{27}$$

and insert it into equation 11:

$$\frac{d^2T}{du^2}+\frac{1}{u}\frac{dT}{du}-T=0 \tag{28}$$

Solutions of this equation are Bessel and Hankel functions with imaginary arguments. To simplify the equations, use the following notations:

$$J_p(x)=\frac{\left(\frac{x}{2}\right)^p}{p!}\sum_{k=0}^{\infty}\frac{p!}{k!(p+k)!}\left(\frac{ix}{2}\right)^{2k} \tag{29}$$

$$H_p^{(1)}(x)=\frac{i}{\sin px}[e^{-p\pi i}J_p(x)-J_{-p}(x)] \tag{30}$$

$$I_0(x)=J_0(ix) \tag{31}$$

$$K_0(x)=H_0^{(1)}(ix) \tag{32}$$

$$I_1(x)=-iJ_1(ix) \tag{33}$$

$$K_1(x)=-iH_1^{(1)}(ix) \tag{34}$$

$$\frac{dI_0(x)}{dx}=I_1(x) \tag{35}$$

$$\frac{dK_0(x)}{dx}=K_1(x) \tag{36}$$

The general solution of equation 28 is

$$T=\theta_1I_0\left(\frac{r}{R}\right)+\theta_2K_0\left(\frac{r}{R}\right) \tag{37}$$

Substitute in equation 5

$$q=-\frac{2\pi rsk}{R}\left[\theta_1I_1\left(\frac{r}{R}\right)+\theta_2K_1\left(\frac{r}{R}\right)\right] \tag{38}$$

The integration constants θ_1 and θ_2 are determined by the boundary conditions

$r=r_0$: $T=T_0$ and $q=q_0$

$r=r_1$: $q=0$

Hence, the following is obtained:

$$\theta_1=\frac{T_0K_1\left(\frac{r_1}{R}\right)}{I_0\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)-I_1\left(\frac{r_1}{R}\right)K_0\left(\frac{r_0}{R}\right)} \tag{39}$$

$$\theta_2=\frac{T_0I_1\left(\frac{r_1}{R}\right)}{K_0\left(\frac{r_0}{R}\right)I_1\left(\frac{r_1}{R}\right)-K_1\left(\frac{r_1}{R}\right)I_0\left(\frac{r_0}{R}\right)} \tag{40}$$

Inserting these constants into equations 37 and 38 yields

$$T=T_0\left[\frac{I_0\left(\frac{r}{R}\right)K_1\left(\frac{r_1}{R}\right)-K_0\left(\frac{r}{R}\right)I_1\left(\frac{r_1}{R}\right)}{I_0\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)-K_0\left(\frac{r_0}{R}\right)I_1\left(\frac{r_1}{R}\right)}\right] \tag{41}$$

$$q=4\pi hrRT_0\times\left[\frac{I_1\left(\frac{r_1}{R}\right)K_1\left(\frac{r}{R}\right)-I_1\left(\frac{r}{R}\right)K_1\left(\frac{r_1}{R}\right)}{I_0\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)-I_1\left(\frac{r_1}{R}\right)K_0\left(\frac{r_0}{R}\right)}\right] \tag{42}$$

and the initial heat flow into the fin

$$q_0=4\pi hrRT_0\times\left[\frac{I_1\left(\frac{r_1}{R}\right)K_1\left(\frac{r_0}{R}\right)-I_1\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)}{I_0\left(\frac{r_0}{R}\right)K_1\left(\frac{r_1}{R}\right)-I_1\left(\frac{r_1}{R}\right)K_0\left(\frac{r_0}{R}\right)}\right] \tag{43}$$

References

1. MATHEMATICAL EQUATIONS FOR HEAT CONDUCTION IN THE FINS OF AIR COOLED ENGINES, D. R. Harper III, W. B. Brown. *Report no. 158*, National Advisory Committee for Aeronautics, Washington, D. C., May 11, 1922, pp. 3–32.

2. EFFICIENCY OF EXTENDED SURFACES, K. A. Gardner. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 67, Nov. 1945, pp. 621–31.

3. HEAT TRANSFER (book), Max Jakob. John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 229–38.

4. HEAT TRANSMISSION (book), William H. McAdams. McGraw-Hill Book Company, Inc., New York, N. Y., 1942.

5. DESIGN OF FINS FOR COOLING OF SEMICONDUCTORS, Werner Luft. *Electrical Manufacturing*, New York, N. Y., vol. 60, Nov. 1957, pp. 98–103.

6. TABLES OF FUNCTIONS (book), Eugene Jahnke, Fritz Emde. Dover Publications, New York, N. Y., 1945, pp. 126–268.

◆

Discussion

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa.): The curves presented in this paper should be very helpful in designing cooling fins in which the heat flow is generally radial. A fuller treatment of forced air cooling, even though much less rigorous, would make the paper more useful. Since the heat transfer to air controls the "natural radius," a great deal of work must be done before the curves given in the paper can be used. This is not the fault of the author, but of the existing literature.

The authors undoubtedly had good reason for using the meter instead of the centimeter as the unit of length. Having chosen the meter, however, its use should have been made consistent throughout the paper to avoid confusion. Special care must be taken in applying heat transfer data from other sources to equations in Appendix II.

In their conclusion, the authors emphasize the fact that extended cooling surfaces consisting of many narrow fins require less material than one large fin, although other economic factors will influence the final design. The authors mention that one of the objections to multiple fins is the amount of material required at the hub to transmit heat from the semiconductor to the multiple fins. It should be noted that this mass of material need not be wasted if it is used both for heat transmission to the fins and as the electric bus. In this arrangement the fins are attached to the sides of a bus bar on which all parallel cells are mounted. This has permitted us to take full advantage of paralleling busses, voltage divider transformer circuits, and compact current-balancing reactors. A large forced air-cooled semiconductor rectifier of this design has been found economical in material, labor, and space.

Forward Current Surge Failure in Semiconductor Rectifiers

F. E. GENTRY
ASSOCIATE MEMBER AIEE

THE APPLICATION of semiconductor rectifiers to most practical circuits requires a knowledge of the device's ability to withstand the forward current surges encountered as a result of severe faults or high instantaneous overloads. In particular, an understanding of the nature of failure is desirable in order to properly protect the rectifier. Suitable fuses can then be chosen to obtain optimum utilization of the rectifier. Gutzwiller[1] has developed fuse-selection techniques which are adequate for this purpose.

The purpose of this paper is to establish a better understanding of forward current surge failure. In particular, surge failure as a function of current and time is investigated for times less than 10 milliseconds.

Mechanism of Failure

Forward current surge failure can usually be attributed to melting of the semiconductor or the materials to which it is attached. This occurs when the semiconductor reaches high enough temperature for the junction to melt and thus be damaged or destroyed. Fig. 1 shows a microsectioned[2] junction which melted as a result of the application of a half-sine-wave surge of 8.3-millisecond duration. In this case both the base contact and junction solders alloyed with the silicon to form a metallic short circuit through the rectifying junction.

Both forward and reverse current contribute to the temperature rise. The reverse heating is indirect in that the forward surge must raise the semiconducting material to a temperature sufficiently high for the reverse power to make a significant contribution to the total heat generated. When this contribution is such that the temperature continues to rise after the application of reverse voltage, the unit will thermally run away. This sequence is expected since reverse current increases exponentially with increasing temperature. The magnitude of the heat generated is determined by the forward and reverse voltage-current characteristic, the junction temperature, reverse voltage, junction area, surge current, and the thermal properties of the semiconductor and adjoining materials.

In addition to failure because of melting two other types of failure occur occasionally. In particular, if the semiconductor is prestressed as a consequence of poor design or fabrication techniques, the additional thermal stresses resulting from the high-temperature gradients caused by the surge may result in fracture. Proper mounting to minimize stress has been discussed by Henkels.[3] Severe damage which cannot be explained by either melting or fracture sometimes occurs because of breakdown on the surface of the semiconductor. It is suspected that this type of failure is caused by contamination, since these units can usually be recovered by suitable surface cleaning.

Nomenclature

T = temperature rise above steady-state temperature, C (degrees centigrade)
T_M = temperature necessary to melt the rectifying junction or cause transient thermal runaway, C
T_J = steady-state junction temperature, C
t = duration of surge, seconds
x = distance from junction, inches
k = thermal conductivity, (watt-inch)/(inch^2 C)
c_p = specific heat, (watt-second)/(pounds C)
ρ = density, pounds/inch^3
$\alpha = k/c_p$ thermal diffusivity, inch^2/second
$G(t)$ = heat per unit volume, watts/inch^3
l = length of middle zone, inch
R = resistance of middle zone, ohms
I = rms value of current, amperes
V = instantaneous voltage, volts
A_J = area of junction, inches^2
v = effective volume of semiconductor, inch^3
i = instantaneous current, amperes

subscripts
1 = material 1
2 = material 2

Analysis of Failure

MELTING

For melting to occur the semiconductor must reach a temperature which will cause it or the adjacent materials to melt and damage or destroy the junction. Consequently, the problem becomes one of determining how this temperature is related to current, time, and the device parameters. Diebold[4] has considered this problem for the case where heat is generated only at the junction. For current densities in excess of 1,000 amperes/centimeter^2 an appreciable voltage drop also occurs across the middle zone of the rectifier. Kleinman,[5] Hall,[6] and Sah, Noyce, and Shockey[7] have predicted this drop. However, the exact equations will be approximated because their complexity prohibits their use in this analysis.

Using the model shown in Fig. 2, the voltage drop is considered to consist of three parts:

V_J = voltage drop across the junction
V_m = voltage drop across the middle zone of the rectifier
V_e = voltage drop caused by the resistance of the materials not included in regions *1* and *2*

V_e is subtracted for this analysis since the heat produced in the materials other than the semiconductor and its junction is

---

Paper 58-927, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 13, 1958; made available for printing April 23, 1958.

F. E. GENTRY is with the General Electric Company, Clyde, N. Y.

This paper is an extract from a thesis to be submitted at Syracuse University in partial fulfillment of the requirements for a master's degree in Electrical Engineering.

Vo
Pt.
19

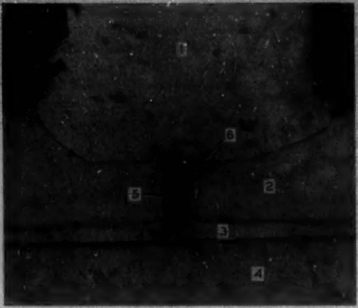


Fig. 1 (left). A microsectioned silicon rectifier which failed as result of forward current surge

1—Aluminum wire
2—Silicon
3—Gold antimony solder
4—Kovar counterelectrode
5—Portion of rectifier which melted
6—P-N junction

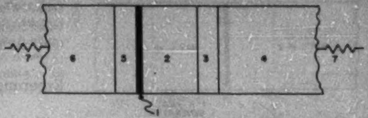


Fig. 2. One-dimensional model of a semiconductor rectifier

1—P-N junction
2—Middle zone of semiconductor
3—Base solder
4—Base counterelectrode
5—Junction solder
6—Junction counterelectrode
7—Ohmic resistance (from counterelectrodes to external connections)

relatively small and has little effect on the temperature rise of the junction. The drop across the semiconductor then becomes V_J+V_m.

As a first approximation the resistance of the middle zone is considered to be constant. The voltage drop across the middle zone is then iR, with the remainder of the drop occurring at the junction. The total drop then becomes V_J+iR. If this equation is fitted to a typical rectifier curve as shown in Fig. 3, a good approximation is realized if V_J is assumed to be constant. The heat generated then becomes $Vi=V_Ji+i^2R$, where V_Ji is the heat generated at the junction and i^2R is the heat generated in the middle zone.

For very short times the model shown in Fig. 4 can be used. The peak junction temperature rise for a rectangular current pulse is then determined and set equal to T_M; see Appendix I. This leads to the equation

$$\frac{R\sqrt{\alpha_1}}{l_1}(I^2t)+\frac{2V_J}{\sqrt{\pi}}(\sqrt{I^2t})-A_J(T_M-T_J)\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)=0 \quad (10)$$

Failure will occur when the junction reaches the destructive temperature T_M. Examination of the equation reveals that for a fixed set of rectifier parameters, a solution is $I^2t=C_1$, a constant. This means that for failure to occur the product I^2t must exceed C_1. To obtain an insight into the effect the current waveshape would have on the solution, a similar analysis was made using a ramp function for the current waveshape. This solution was found to be

$$\frac{R\sqrt{\alpha_1}}{l_1}(I^2t)+\frac{4V_J}{\sqrt{3\pi}}(\sqrt{I^2t})-A_J(T_M-T_J)\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)=0$$

Further calculations (see Appendix II) show these solutions to be reasonable for $t<l^2/\alpha$. For typical silicon and germanium rectifiers now in production the time l^2/α is:

$l^2/\alpha=500$ microseconds for silicon
$l^2/\alpha=4$ milliseconds for germanium

Using equation 10, calculations are compared with experimental results for a silicon rectifier; see Fig. 5 and Appendix III.

For longer times a distributed constant model becomes rather difficult to analyze and it is convenient to use a model with lumped parameters, as shown in Fig. 6, for further study. For times longer than 10 milliseconds most surge currents are usually repetitive pulses. As a result, analytical study is considerably more difficult. However, the problem is particularly adaptable to solution using an electrical analog of the thermal circuit. The effect of reverse heating, change in characteristics, and complex waveshapes could also be considered by using feedback in the analog.

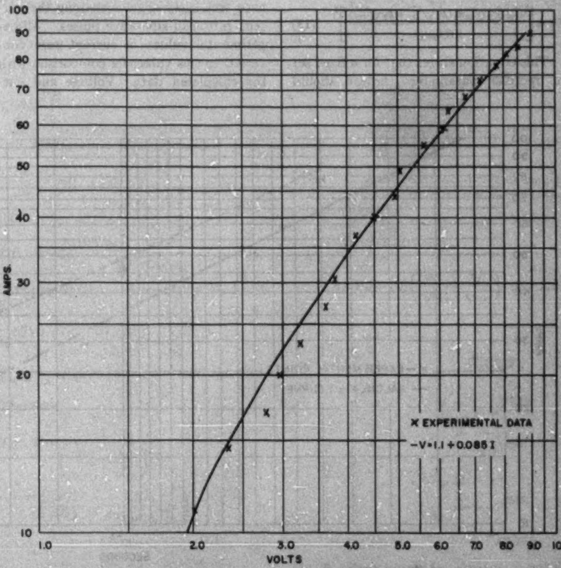


Fig. 3. Forward voltage-current curve for silicon rectifier with area of 5×10⁻⁴ inch²

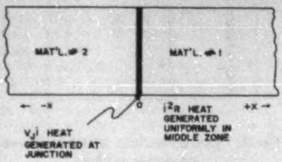

MAT'L. # 2
MAT'L. # 1
← −x
0
+x →
$V_J i$ HEAT GENERATED AT JUNCTION
i^2R HEAT GENERATED UNIFORMLY IN MIDDLE ZONE

Fig. 4. Model used for mathematical analysis

An analytical study using a more refined model, in which a finite solid with internal heat generation and heat generated at one face is bounded on each side by an infinite slab, has been conducted. However, this solution is not included in this paper.

Fracture

For fracture to occur the semiconductor material must be strained beyond its yield point. If the built-in stresses plus the stresses set up by the temperature gradient resulting from the surge reach a critical value S_c, the yield point will be exceeded and the semiconductor will fracture. The temperature gradient is determined by differentiating equation 6 and evaluating it at $x=0$; see Appendix IV. The result is

$$I^4t \approx \left[\frac{C_3 l A_J \sqrt{\pi} k_1 (k_2\sqrt{\alpha_1}+k_1\sqrt{\alpha_2})}{2Rk_2\alpha_1}\right]^2 \qquad (15)$$

This equation states that for a fixed set of rectifier parameters, failure should occur when the product I^4t exceeds the value found by solving the foregoing equation for I^4t.

Experimental Procedure

The units used to establish the experimental points in Fig. 5 were selected from the same production lot and were matched in the forward direction to assure uniform results. Their junction area was approximately 5×10^{-4} inch². The tests shown in Fig. 5 were conducted at 25 C with no load and no reverse voltage applied in order to simulate the conditions used in the mathematical analysis. After each surge the units were checked for reverse characteristic degradation and were considered to be failures if the reverse current exceeded 5 microamperes at rated voltage. In general, the units showed no degradation until the surge current was within 10% of that necessary to cause catastrophic failure.

The basic circuits for generation of the current pulses are shown in Figs. 7 and 8. The circuit shown in Fig. 7 permits the application of steady-state load current and reverse voltage when needed. In addition, any desired number of repetitive pulses can be obtained with a suitable timer. For shorter time pulses the circuit shown in Fig. 8 was used. The inductance was necessary for shaping the current into half-sine-wave pulses. All experimental values of current were converted to rms values for comparison with the calculated data. Voltage and current measurements during the surge were made using peak-reading voltmeters with a response time of 2 microseconds.

The experimental points for the forward voltage-current curve shown in Fig. 3 were taken using a 300-microsecond current pulse. The values indicated are peak values.

General Observations

Surge failure in germanium and silicon rectifiers normally results in the junction being short-circuited. However, if the the magnitude or duration of the surge current becomes several times greater than that of the failure point, vaporization of materials may occur, causing the unit to open the circuit or even explode. For times less than 8 milliseconds, most germanium and silicon rectifiers tested by the author fail when the product $I^2t = C_1$ is exceeded, where C_1 is a constant determined by rectifier design and steady-state junction temperature. With presently available designs the temperature rise necessary to cause failure is somewhat lower for germanium than for silicon rectifiers. Consequently the effect of initial junction temperature is much more pronounced for germanium than for silicon devices.

For germanium and silicon rectifiers, the resistive portion of the forward characteristic increases with increasing temperature. When a surge of forward current is applied the heat generated in the middle zone increases as the tempera-

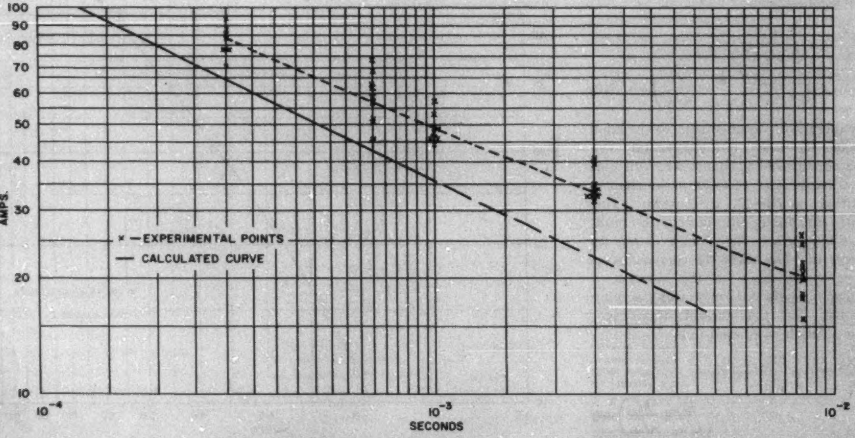


Fig. 5. Surge-current failure curve for silicon rectifier with junction area of 5×10^{-4} inch²

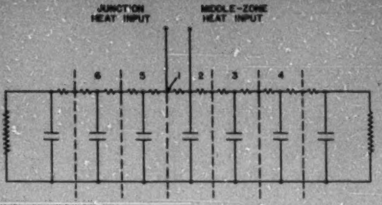


Fig. 6 (left). Equivalent thermal circuit for semiconductor rectifier. Each material is represented by one or more T-networks representing its thermal resistance and capacitance

1—P-N junction
2—Semiconductor
3—Base solder
4—Base counterelectrode
5—Junction solder
6—Junction counterelectrode

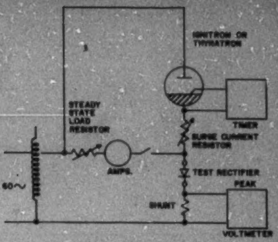


Fig. 7. Basic circuit for generating one or more current surges of 8.3-millisecond duration

ture of the semiconductor rises. This phenomenon is probably a controlling factor in determining the surge failure point.

Conclusions

Surge current failure in germanium and silicon rectifiers is usually the result of very high temperature rise within the semiconductor. Severe damage of the junction then causes the unit to short-circuit itself by localized melting of the semiconductor or its adjacent contacts.

The failure point is dependent on the rectifier design, initial junction temperature, reverse voltage, and the magnitude, shape, and duration of the applied surge current. In general, for surges of less than 1-millisecond duration, germanium and silicon rectifiers fail when the product I^2t exceeds a constant value dependent on the design and characteristics. For longer periods of time materials other than those immediately adjacent to the semiconductor become significant and alter the surge characteristics.

Failure due to fracture should occur when the product I^2t exceeds a constant value dependent on the design and characteristics. However, this mechanism of failure was not observed in the rectifiers tested.

Appendix I

Consider the two adjacent semi-infinite solids shown in Fig. 4. The equation[8] describing the flow of heat in these materials is

$$\frac{\partial T(t,x)}{\partial t}=\alpha\frac{\partial^2 T(t,x)}{\partial x^2}+\frac{G(t)}{\rho c_p} \tag{1}$$

with the initial and boundary conditions

$$T_1(0,x)=T_2(0,x)=\frac{\partial T_1(t,\infty)}{\partial x}=\frac{\partial T_2(t,-\infty)}{\partial x}=0$$

$$k_2\frac{\partial T_2(t,0)}{\partial x}-k_1\frac{\partial T_1(t,0)}{\partial x}=\frac{IV_J}{A_J}$$

and the assumptions

$$G_1(t)=\frac{I^2R}{v_1}$$

$$G_2(t)=0$$

Solving by Laplace transforms gives, for material 2,

$$\frac{\partial^2 T_2(s,x)}{\partial x^2}-\frac{s}{\alpha_2}T_2(s,x)+\frac{T_2(0,x)}{\alpha_2}+\frac{G_2(s)}{\rho_2 c_{p2}}=0 \tag{2}$$

Since

$$G_2(s)=0$$

$$T_2(0,x)=0$$

$$\frac{\partial T_2(s,-\infty)}{\partial x}=0$$

then

$$T_2(s,x)=B\epsilon^{\sqrt{s/\alpha_2}\,x} \tag{3}$$

For material 1,

$$\frac{\partial^2 T_1(s,x)}{\partial x^2}-\frac{sT_1(s,x)}{\alpha_1}+\frac{T_1(0,x)}{\alpha_1}+\frac{G_1(s)}{\alpha_1\rho_1 c_{p1}}=0 \tag{4}$$

Since

$$G_1(s)=\frac{I^2R}{sv_1}$$

$$T_1(0,x)=0$$

$$\frac{\partial T_1(s,\infty)}{\partial x}=0$$

then

$$T_1(s,x)=C\epsilon^{-\sqrt{s/\alpha_1}\,x}+\frac{I^2R}{s^2\rho_1 c_{p1}v_1} \tag{5}$$

At the boundary between materials 1 and 2

$$k_2\frac{\partial T_2(s,0)}{\partial x}-k_1\frac{\partial T_1(s,0)}{\partial x}=\frac{V_JI}{sA_J}$$

and $T_1(s,0)=T_2(s,0)$

Therefore

$$T_1(s,x)=\frac{1}{\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times\left(\frac{V_JI}{s^{3/2}A_J}-\frac{I^2Rk_2}{s^2\rho_1c_{p1}v_1\sqrt{\alpha_2}}\right)\epsilon^{-\sqrt{s/\alpha_1}\,x}+\frac{I^2R}{s^2\rho_1c_{p1}v_1} \tag{6}$$

Transforming to a function of t gives

$$T_1(t,x)=\frac{1}{\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\frac{V_JI}{A_J}\times\left(2\sqrt{\frac{t}{\pi}}\,\epsilon^{-\frac{x^2}{4\alpha_1 t}}-\frac{x}{\sqrt{\alpha_1}}\operatorname{erfc}\frac{x}{2\sqrt{\alpha_1 t}}\right)-\frac{I^2Rk_2}{\sqrt{\alpha_2}\rho_1c_{p1}v_1}\int_0^t\operatorname{erfc}\frac{x}{2\sqrt{\alpha_1\tau}}\,d\tau+\frac{I^2Rt}{\rho_1c_{p1}v_1} \tag{7}$$

At the interface $x=0$ the temperature is

$$T_1(t,0)=\frac{1}{\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times\left(\frac{2V_J}{A_J\sqrt{\pi}}\sqrt{I^2t}+\frac{Rk_1I^2t}{\rho_1c_{p1}v_1\sqrt{\alpha_1}}\right) \tag{8}$$

or

$$T_1(t,0)=\frac{1}{A_J\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times\left(\frac{2V_J\sqrt{I^2t}}{\sqrt{\pi}}+\frac{R\sqrt{\alpha_1}I^2t}{l_1}\right) \tag{9}$$

where R/l_1 is the resistance per unit length in material 1. When $T_1(t,0)=T_m-T_J$, the equation becomes

$$\frac{R\sqrt{\alpha_1}}{l_1}(I^2t)+\frac{2V_J}{\sqrt{\pi}}(\sqrt{I^2t})-A_J(T_M-T_J)\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)=0 \tag{10}$$

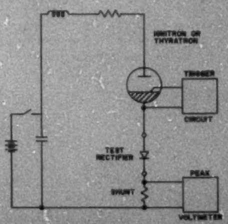


Fig. 8. Basic circuit for generating current surges of less than 5-millisecond duration

Appendix II

To determine the period of time for which the semi-infinite solution is valid, the solution for constant heat into the face of a finite slab is compared with that for a semi-infinite slab. These two solutions are developed by Carslaw and Jaeger.[8] At the face of the slabs the temperatures are: for the semi-infinite slab,

$$T_I=\frac{VI}{k}\left(\frac{2\sqrt{\alpha t}}{\sqrt{\pi}}\right)$$

for the finite slab,

$$T_F=\frac{VIl}{k}\left[1-\frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{\epsilon^{-\frac{\alpha\pi^2(2n+1)^2t}{4l^2}}}{(2n+1)^2}\right]$$

The ratio T_I/T_F is shown in Fig. 9. It is evident that the semi-infinite solution is a reasonably good approximation to the finite solution until t approaches l^2/α.

Appendix III

Using equation 10 and rectifier parameters taken from one of the silicon rectifiers used to establish the experimental characteristic curve in Fig. 3, a theoretical surge-failure curve can be predicted (see Fig. 5). The temperature T_M was chosen to be 660 C since at this temperature the aluminum adjoining the junction should melt sufficiently to cause damage. The following values were used to establish the calculated curve:

$V_J=1.1$ volts
$R=0.085$ ohm
$\alpha_1=0.073$ inch²/second
$\alpha_2=0.14$ inch²/second
$k_1=2.1$ (watt-inch)/(inch² C)
$k_2=5.6$ (watt-inch)/(inch² C)
$A_J=5.0\times10^{-4}$ inch²
$T_J=25$ C
$T_M=660$ C
$l=0.005$ inch

Substituting these values into equation 10 gives

$$4.6I^4t+1.24\sqrt{I^4t}-7.24=0$$

and solving for I^4t gives

$$I^4t=1.27$$

This value is somewhat lower than that obtained by experiment. This discrepancy can be accounted for, since the rectifiers used to establish the experimental curve were not one dimensional and diffusion of heat into other areas of the silicon pellet would occur.

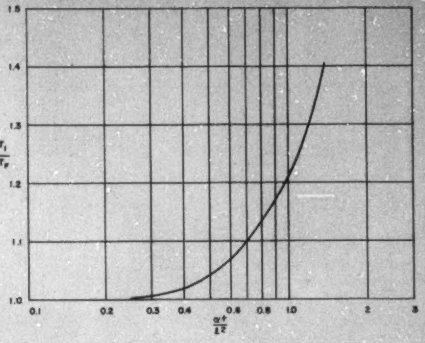


Fig. 9. Ratio of temperature rises for constant heat into face of infinite and finite slabs

Appendix IV

To determine how current and time affect the stresses in the semiconductor the temperature gradient must be obtained. Differentiating equation 6 gives

$$\frac{\partial T(s,x)}{\partial x}=\frac{1}{\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times \left(\frac{-V_JI}{sA_J\sqrt{\alpha_1}}+\frac{I^2Rk_2}{s^{3/2}\rho_1c_{p1}v_1\sqrt{\alpha_1\alpha_2}}\right)\epsilon^{-\sqrt{s/\alpha_1}\,x} \tag{11}$$

Taking the inverse transform gives

$$\frac{\partial T_1(t,x)}{\partial x}=\frac{1}{\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times \left[-\frac{V_JI}{A_J\sqrt{\alpha_1}}\operatorname{erfc}\frac{x}{2\sqrt{\alpha_1t}}+\frac{2I^2Rk_2}{\rho_1c_{p1}v_1\sqrt{\alpha_1\alpha_2\pi}}\left(\sqrt{t}\epsilon^{-\frac{x^2}{4\alpha_1t}}-\frac{x}{\sqrt{\alpha_1}}\operatorname{erfc}\frac{x}{2\sqrt{\alpha_1t}}\right)\right] \tag{12}$$

At $x=0$

$$\frac{\partial T(t,0)}{\partial x}=-\frac{1}{A_J\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times \left(-\frac{V_JI}{\sqrt{\alpha_1}}+\frac{2R}{l\sqrt{\pi}}\frac{k_2\sqrt{\alpha_1}}{k_1\sqrt{\alpha_2}}\sqrt{I^4t}\right) \tag{13}$$

When

$$\frac{\partial T_1(t,0)}{\partial x}=C_2$$

where C_2 is the temperature gradient which results in the yield point being exceeded, then

$$C_2\approx\frac{2R}{lA_J\sqrt{\pi}}\times\frac{k_2\sqrt{\alpha_1}}{k_1\sqrt{\alpha_2}\left(\frac{k_2}{\sqrt{\alpha_2}}+\frac{k_1}{\sqrt{\alpha_1}}\right)}\times\sqrt{I^4t} \tag{14}$$

or

$$I^4t\approx\left[\frac{C_2lA_J\sqrt{\pi}k_1(k_2\sqrt{\alpha_1}+k_1\sqrt{\alpha_2})}{2Rk_2\alpha_1}\right]^2 \tag{15}$$

References

1. The Current-Limiting Fuse as Fault Protection for Semiconductor Rectifiers, F. W. Gutzwiller. *AIEE Transactions*, vol. 77, pt. I, Nov. 1958, pp. 751–55.

2. Microsectioning: A Metallographic Technique for Semiconductor Devices, J. S. Hanson. *Journal*, International Business Macnines Corporation, New York, N. Y., July 1957, pp. 279–88.

3. The Fused Silicon Rectifier, Herbert W. Henkels. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 733–46.

4. Temperature Rise of Solid Junctions Under Pulse Load, E. J. Diebold. *Ibid.*, vol. 76, pt. I, Nov. 1957, pp. 593–98.

5. The Forward Characteristic of the PIN Diode, D. A. Kleinman. *Bell System Technical Journal*, New York, N. Y., vol. 35, no. 3, May 1956, pp. 685–706.

6. Power Rectifiers and Transistors, R. N. Hall. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, no. 11, Nov. 1952, pp. 1512–18.

7. Carrier Generation and Recombination in PN Junctions and PN Junction Characteristics, C. T. Sah, R. N. Noyce, W. Shockley. *Ibid.*, vol. 45, no. 9, Sept. 1957, pp. 1228–43.

8. Conduction of Heat in Solids (book), H. S. Carslaw, J. C. Jaeger. Oxford University Press, London, England, 1947.

See Discussion on Page 754

The Current-Limiting Fuse as Fault Protection for Semiconductor Rectifiers

F. W. GUTZWILLER
ASSOCIATE MEMBER AIEE

THE RAPID DEVELOPMENT of germanium and silicon power rectifier components in recent years has posed unique problems in fault current protection. These two types of rectifiers are characterized by the extremely high current densities tolerable in the active rectifying element at normal loads while maintaining low forward-voltage drop and thus high efficiency. For comparison purposes, full-load current densities (amperes per unit, cross-sectional area) in germanium and silicon rectifiers are approximately 1,000 times greater than in selenium and copper oxide metallic rectifiers.

The Semiconductor Rectifier Under Fault Conditions

The advantage of high continuous current ratings in a small package produces a handicapping corollary: limited thermal capacity of the rectifying element, the semiconductor p-n junction. When a constant load current is applied to a semiconductor rectifier, the temperature of the junction rises and stabilizes within a matter of cycles (at most, perhaps, seconds) at essentially a fixed temperature differential above the temperature of the heat sink to which the junction is attached.

In a similar manner, junction temperatures climb very quickly when overloads are impressed on the cell. Under fault conditions during which currents ranging from a few times to several hundred times normal load values may be expected, junction temperatures may skyrocket to destructive values within a few milliseconds or less. Immediate cell failure will occur as materials in and surrounding the junction melt under intensive local heating.[1] To prevent cell failure under these conditions, either the fault currents must be interrupted very quickly or rectifier components must be drastically derated.

The Current-Limiting Fuse

The current-limiting fuse, developed initially for protection of low-voltage branch feeder circuits connected to stiff systems, is admirably suited for interrupting potentially high fault currents quickly. From a thermal standpoint, current-limiting fuse construction is very similar to a germanium or silicon rectifier. The typical current-limiting fuse consists essentially of a fusible element of high current density, with low thermal capacity connected thermally and electrically to a heat sink of lower current density. This similarity in thermal construction between the current-limiting fuse and the semiconductor rectifier leads to similar thermal characteristics, and thus similar reactions to fault currents up to the point of melting in the case of the fuse and failure in the case of the rectifier cell.

In this paper, the term "current-limiting fuse" applies to a fuse which, when used within its current-limiting ratings, limits the peak let-through current to a value lower than that which would otherwise flow in the circuit, including the impedance of the fuse. It will also limit the clearing time to less than the duration of the first major current loop. The action of a typical current-limiting fuse is indicated in Fig. 1. The dashed line indicates the manner in which fault current would develop without the limiting action of the current-limiting fuse. If permitted to flow in a typical semiconductor rectifier circuit, the rectifier cells might be destroyed before the current reached zero the first time. However, at some point *A*, the current-limiting fuse link melts and the fuse commences to arc. Depending on the fuse design and the external circuit, the current may continue to rise somewhat further to point *B*, the peak let-through current. Beyond this point, the increasing impedance of the arcing fuse forces the fault current down to zero at some point *C*.

The industry refers to the region from fault initiation to *A* as melting time, and from *A* to *C* as arcing time. The sum of melting and arcing times (initiation to point *C*) is designated as clearing time.

This discussion will be limited to time intervals in this region below approximately one cycle where current-limiting fuses are effective. Protection for these time intervals is essential whenever d-c bus faults or cell failures can occur on relatively stiff electrical systems and when good rectifier cells are not considered expendable under these circumstances. In general, systems may be considered stiff for these purposes when the system impedance permits fault currents in excess of seven to ten times normal load current to flow. For faults and overloads at lower current levels, protection which acts more slowly (conventional fuses and circuit breakers) may generally be used.

Co-ordinating Rectifier Cell and Fuse

As shown by Gentry,[1] the forward-current capability of semiconductor power rectifiers in the subcycle region assumes an essentially constant $\int i^2dt$ characteristic at times under a critical value, which depends on the specific cell design. At times over this critical value, the constant minimum $\int i^2dt$ or I^2t value is conservative. (I=rms amperes, t=seconds.)

Current-limiting fuses exhibit a similar constant I^2t relationship. At less than approximately 0.01 second, current-limiting fuses melt at essentially constant I^2t.[2] Fuses may also be rated in terms of maximum clearing I^2t.[3]

Unfortunately, maximum clearing I^2t ratings are available on only a very limited number of types of current-limiting fuses at the time of this writing.[3] Until the

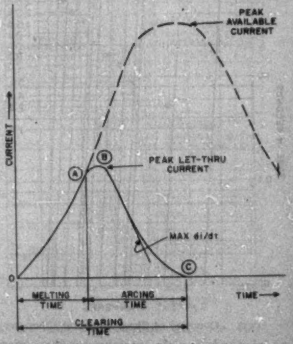


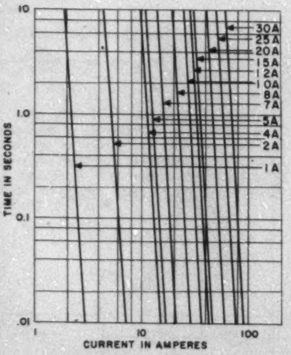
Fig. 1. Limiting action of current-limiting fuse

Paper 58-928, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 23, 1958.

F. W. GUTZWILLER is with the General Electric Company, Clyde, N. Y.

fuse manufacturers make maximum clearing I^2t data available on all types of low-voltage current-limiting fuses, approximations of this value are necessary. Of three fuse manufacturers consulted, all agreed that arcing I^2t should not exceed twice the melting I^2t for properly designed and functioning fuses. Since clearing I^2t is the sum of melting and arcing I^2t, maximum clearing I^2t may be conservatively approximated by multiplying the maximum melting I^2t by three.

Specific melting I^2t ratings as such are not published for most commercial low-voltage current-limiting fuses. Another approximation is therefore necessary until adequate fuse data are made available by manufacturers for this purpose. By referring to the melting-time–current characteristics of a particular fuse, the melting I^2t may be approximated by squaring the melting-current amperes at 0.01 second and multiplying by 0.01 second. This yields a value of ampere^2-seconds. The manufacturer's maximum tolerance in current or time should be introduced into this calculation to produce a more realistic maximum value. For example, Fig. 2 illustrates the melting-time–current characteristics of a typical line of current-limiting fuses of a type developed especially for semiconductor rectifier protection. The maximum rms melting current for the 15-ampere fuse at 0.01 second is 46 amperes plus 5% tolerance, or 48.3 amperes. The maximum melting I^2t approximation is $(48.3)^2 \times 0.01$ second, or 23.3 ampere^2-seconds. The approximation for maximum clearing I^2t of the fuse is 3×23.3 or 69.9 ampere^2-seconds. Actual tests on this particular fuse and similar types of fuses indicate that this method of approximation is very conservative.



Courtesy of Chase-Shawmut Company

Fig. 2. Melting-time–current characteristics for a line of 250-volt current-limiting fuses

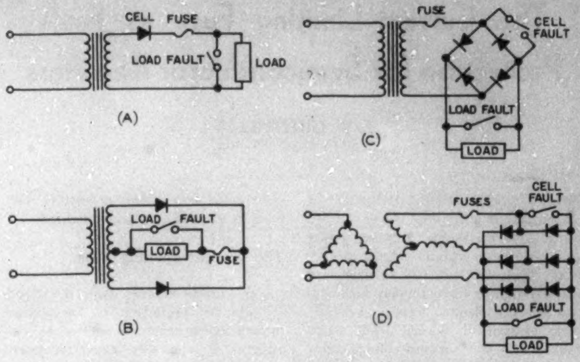


Fig. 3. Some simple current-limiting fuse protection circuits

A—Half wave
B—Full-wave centertap
C—Single-phase double-way or bridge
D—Three-phase double-way or bridge

In its simplest form, protection is provided as shown in Fig. 3(A), a fuse in series with a semiconductor rectifier. Protection of the rectifier will occur for any subcycle waveshape and magnitude of current within the interrupting rating of the current-limiting fuse if the maximum clearing I^2t of the fuse is equal to or less than the minimum I^2t rating of the rectifier cell. Usually the I^2t rating will be specified by the cell manufacturer as a value imposed with the cell already operating at maximum rated junction temperature and peak inverse voltage. Higher I^2t ratings may be available for such circumstances as a lower operating junction temperature preceding the surge (particularly applicable to germanium devices) and conditions where inverse voltage does not follow the fault current, for example, when a fuse interrupts the voltage supply.

The basic protection circuit shown in Fig. 3(A) and the co-ordination principle just described applies to any rectifier circuit in which current-limiting action is desired, and in which the fault current flows in one rectifier cell path at a time. Thus, fuses in the circuits of Fig. 3(B), (C), and (D) should also be selected to have a maximum clearing I^2t lower than the I^2t rating of any single rectifier cell in the circuit.

Line fusing as shown in Fig. 3(A), (C), and (D) protects the rectifier cells against severe d-c short circuits and also protects the remaining good cells against faults that result from a defective cell. The continuous rating of the fuse in any circuit should be adequate for handling maximum rated load indefinitely without blowing. Also, its interrupting rating must be greater than the available fault current in the circuit. Again, it should be emphasized that current-limiting fuses will usually provide positive protection only against severe faults. Other types of protective devices in the form of circuit breakers or conventional fuses should be provided for lower current overloads which the rectifier cells are capable of sustaining for more than a few cycles.

When N rectifier cells are connected in parallel with one another but in series with a single fuse, as shown in Fig. 4, the clearing I^2t of the fuse should be no greater than N^2 times the I^2t rating of the individual rectifier cells, if all the parallel cells share current equally. If the rectifier cells are derated for parallel operation to compensate for inequalities in current sharing, the relationship between the maximum clearing I^2t of the fuse and the I^2t rating of the individual rectifier cell may be expressed as follows (see the Appendix).

$$I^2t_f \le I^2t_c[N(1-S)+S]^2 \qquad (1)$$

where

I^2t_f = maximum clearing I^2t of fuse
$I^2t_c = I^2t$ rating of single rectifier cell
N = number of parallel rectifier cells
S = derating factor for parallel cell operation

When a current-limiting fuse in the primary of a transformer must be co-ordinated with a rectifier in the secondary,

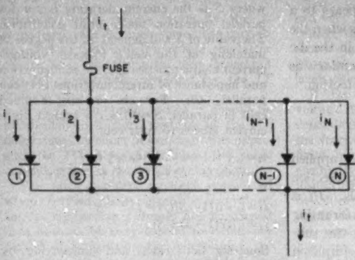


Fig. 4. Co-ordination of single fuse with parallel rectifier cells

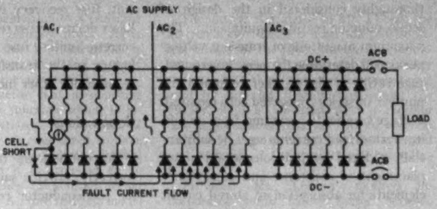


Fig. 5. Three-phase bridge rectifier providing noninterrupted service upon cell failure

their respective I^2t may be compared by the following equation, which is based on the premise that currents vary inversely as the voltage ratio of the windings:

$$I^2t_f \leq \left[\frac{V_s}{V_p}\right]^2 \times I^2t_c \qquad (2)$$

where

V_s = rated secondary voltage of transformer
V_p = rated primary voltage of transformer

Care must be taken in applying current-limiting fuses in the primary circuits of transformers to avoid having the inrush magnetizing current blow the fuse. In transformers operating close to saturation flux densities, inrush magnetizing currents may run as high as ten times full-load current.

Fuse Co-ordination for Isolation of Defective Rectifier Cells

For high-current semiconductor rectifier installations in which many cells are connected in parallel in each leg of the circuit, and where continuity of operation is essential, current-limiting fuses are used to isolate short-circuited cells from the circuit. With a defective cell isolated in less than a cycle, the rectifier equipment can continue to supply d-c power and, provided adequate load-carrying capacity remains in the leg, the defective cell, or cells, and fuse may be replaced at the convenience of the operator. A simple 3-phase bridge circuit of this type is shown in Fig. 5.

This circuit consists of six parallel matched cells in each leg of a bridge. A current-limiting fuse is connected in series with each cell. The fuses must have similar resistances in order not to unbalance the current sharing of the parallel rectifier paths. In the event that a rectifier cell short-circuits, a direct short circuit develops between the a-c lines for a part of each cycle. For example, if the indicated cell in phase 1 short-circuits, a fault current will flow as indicated whenever line AC_1 is positive with respect to line AC_3. To isolate the defective cell satisfactorily, fuse 1 must clear the fault current before any of the rectifier cells in phase 2 fail and before any of the fuses in phase 2 reach the melting point. Equation 1 may be used to co-ordinate fuse 1 with the rectifier cells, this being the case of a fuse in series with a group of parallel cells. To meet the fuse selectivity requirement that only the fuse adjacent to the faulty cell blow, equation 1 may be used by substituting the maximum melting I^2t value of the fuses for I^2t_c in the equation. Taking into account inequalities in parallel cell current sharing, and the ratio of clearing I^2t to melting I^2t for typical types of fuses, leads to the general conclusion that no less than three cells should be operated in parallel in a leg of this type of circuit if positive fuse selectivity is to be realized.

In the circuit of Fig. 5, the five parallel cells and fuses remaining in phase 1 after the faulty cell has been isolated should be capable of carrying rated current until replacement of the defective cell has been made.

Where d-c bus faults or overloads may be expected, a circuit breaker should be located in the d-c output of the rectifier to disconnect the fault before rectifier cell failure or fuse melting can occur. Tripping of a-c breakers may also be arranged to perform this function. Low-voltage current-limiting fuses are currently available for d-c bus interruption duty in ratings up to 1,200 amperes at 250 volts. A protective co-ordination chart of the type that has been described in the literature[4] is a useful tool in determining that the d-c fault protection functions before either rectifier cells or isolating fuses reach their respective maximum ratings under any d-c fault condition from a bolted bus fault, if this is a possibility, to mild overloads. An example of such a d-c fault co-ordination chart for a 400-ampere rectifier supply is shown in Fig. 6. For all current levels up to the maximum possible d-c bus fault, the circuit-breaker characteristic lies to the left of both the rectifier cell and fuse characteristics indicating satisfactory co-ordination for d-c faults.

Transient Voltage Considerations

The rectifying characteristics of germanium and silicon cells can be destroyed by voltage in excess of the transient peak inverse voltage rating, even if this excessive voltage is impressed across the cell for a small fraction of a cycle. For this reason, transient voltages must be

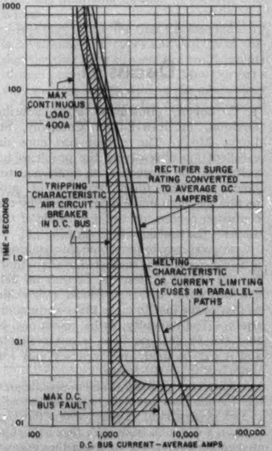


Fig. 6. Fault co-ordination chart for 400-ampere rectifier equipment of type shown in Fig. 5

thoroughly considered in the design of semiconductor rectifier equipment. The maximum magnitude of transient voltage peaks will determine the peak inverse voltage rating of the rectifier cells and the number of series-connected cells required. Voltage transients occurring from switching normal load currents may be substantially reduced by judicious selection and placement of resistance and capacitance elements for absorption of stored circuit energy. However, these techniques have little effect in dampening the transient voltages induced when a fault current with its extremely high associated energy is suddenly interrupted. To minimize induced voltages from fault currents, protective devices should not be over-abrupt in their interrupting action on fault current. In the case of the current-limiting fuse action described in Fig. 1, the maximum indicated slope $(-di/dt)$ of the current during arcing is an indication of the recovery voltage developed across the fuse. The magnitude of fuse recovery voltage is thus a function of fuse design and the parameters of the circuit in which the fuse is used.

In the parallel cell scheme of Fig. 5, the fuse recovery voltage is impressed directly across the cells in parallel with the fuse. In this type of application, the need for the right combination of fuse characteristics is obvious: the fuse should offer an optimum compromise of low clearing I^2t and low recovery voltage.

Depending on the specific rectifier circuit, fuse recovery voltage appears to a lesser degree across rectifier cells when the current-limiting fuse is located in the d-c bus or in the transformer secondary as required for other modes of protection.

Conclusions

Current-limiting fuses, when applied correctly, provide satisfactory protection for semiconductor rectifier cells. I^2t is proposed as a satisfactory means for rating semiconductor rectifier cells in the subcycle region. Satisfactory and simplified co-ordination between cells and fuses may be achieved by properly relating cell I^2t rating to maximum fuse clearing I^2t. This eliminates the necessity for complex fault current calculations in the subcycle region.

Appendix. Co-ordination of Current-Limiting Fuses with Parallel Rectifier Cells

Co-ordination between parallel cells and a current-limiting fuse is most critical when one rectifier cell (cell 1 in Fig. 4) is carrying maximum rated fault current, and all other parallel cells have maximum parallel mismatch, and are therefore carrying a minimum share of the total fault current i_t. In this figure there are N parallel rectifier paths, and each of the parallel paths other than cell 1 is assumed to be carrying a minimum current equal to

$$i_2=i_3=i_{N-1}=i_N=i_1(1-S)$$

where S is the current derating factor for parallel operation under fault conditions. The value of S will depend on the degree of matching of the cell's forward voltage-current characteristics and the configuration and impedance of interconnections between parallel cells. For six or less matched cells in parallel, $S=0.1$ for a typical high-current silicon rectifier cell.

$$i_t=i_1+i_2+i_3+\ldots+i_{N-1}+i_N$$
$$i_t=i_1+i_1(1-S)+i_1(1-S)+\ldots i_1(1-S)+i_1(1-S)$$
$$i_t=i_1+i_1(1-S)(N-1)$$
$$i_t=i_1[N(1-S)+S]$$

Squaring both sides, and multiplying by time t yields

$$i_t^2t=i_1^2[N(1-S)+S]^2t$$

By defining all currents as rms values and substituting the I^2t ratings of the fuse and the rectifier cells

$$I^2t_f=I^2t_c[N(1-S)+S]^2 \qquad (1)$$

References

1. See pages 746–50 of this issue.

2. CURRENT-LIMITING FUSES: THEIR CHARACTERISTICS AND APPLICATIONS, **P. C. Jacobs, Jr.** *AIEE Transactions*, vol. 75, pt. III, Oct. 1956, pp. 988–93.

3. HIGH CAPACITY CURRENT LIMITING FUSES TODAY, **E. M. Fitzgerald, V. N. Stewart.** *AIEE CP 58-131* (available on request).

4. SILICON POWER RECTIFIER EQUIPMENT, **J. A. Marshall, V. N. Stewart.** *AIEE Transactions*, vol. 76, pt. II, 1957 (Jan. 1958 section), pp. 366–69.

5. CO-ORDINATED FUSE PROTECTION FOR LOW-VOLTAGE DISTRIBUTION SYSTEMS IN INDUSTRIAL PLANTS, **John C. Lebens.** *Ibid.*, vol. 73, pt. II, May 1954, pp. 77–83.

6. RATING AND APPLICATION OF GERMANIUM AND SILICON RECTIFIERS, **F. W. Gutzwiller.** *Ibid.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 753–57.

Discussion

J. C. Lebens (Bussmann Manufacturing Division, McGraw-Edison Company, St. Louis, Mo.): The author has done an excellent job of resolving the problems involved in the protection of semiconductor junctions. The tough problem of protection has been made more difficult by the confusion concerning the magnitude of the amperes the fuse must interrupt.

One good way to overcome this confusion is by the use of I^2t as suggested by the author. He can obtain by test the I^2t that the cell can withstand and then equate this value to that of the fuse to make his co-ordination study. The difficulty, as he points out, is that the I^2t data on fuses is not available and it can be used with accuracy only for times less than 0.01 second.

In our study of the problem we did not use I^2t but converted the data on the cell to rms symmetrical amperes which put the data on a common basis with that available on commercial fuses. By using the concept of prospective current and virtual time the co-ordination between the safe time of the cell and the total clearing time of the fuse at all values of load may be established.

When a full wave rectifier fails, the current which the undamaged cell and its fuse sees is a pulse which may have different shapes depending on the characteristics of the circuit. Actually our tests indicate that representative data can be obtained by testing a single cell in series with the fuse on a low-voltage a-c circuit.

These tests showed that the damage data on the cell are conservative, once they are properly defined in rms symmetrical amperes. Because of this we are able to produce a current-limiting fuse of simplified construction which matches the damage characteristic of the cell. This substantially reduces the cost of the protection.

Originally we also were quite concerned with the effect of preloading on the capacity of the cell. Actually, preloading does not shift the damage characteristic appreciably.

Fundamentally, for complete protection, the thermal capacity of the fuse should be the same as that of the cell and it should be subjected to the same cooling. Practically this is virtually impossible so that tests were run to determine the effect of preloading and cooling on the cell damage characteristic. We were pleased to find that the effect did not make the co-ordination of the cell and the current-limiting fuse of simplified design impossible. The author has been instrumental in developing and in bringing into proper perspective the complete and accurate data so necessary to the solution of this problem.

Joint Discussion of Paper Nos. 58-927 and 58-928

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa.): This paper and the paper by Gentry[1] attempt to lay the ground work for the selection of current-limiting fuses for the protection of semiconductor rectifiers on the basis of the I^2T capability of the individual cells. Since the forward resistance of semiconductors is far from linear, the use of I^2T as a criterion must be questioned. If the semiconductor had a constant voltage drop independent of current, its capability would be a function of IT. In actual fact the heat generated will fall somewhat between these two limits.

There is danger in the use of I^2T determined by short pulses of high current, when applied to circuits or conditions in which the peak fault current is limited to a considerably lower value. Normally this is considered safe because some heat may be assumed to flow into the sink in the latter case. However, the total heat generated can be considerably increased by the IT component.

If an IT component exists, it should be manifested in a sharp decrease of the I^2T determined by short pulses, as in this paper, and the I^2T determined from the surge current curves in the 1-to-3 cycle region.

The same problem must be considered when expressing fault currents in rms values. In this connection I should like to suggest that authors be very careful to define the period of integration for determining the rms value, especially when the periods are longer than $^1/_2$ cycle and shorter than 2 or 3 cycles. This applies to both cell and fuse manufacturers.

Figs. 5 and 6 refer to a 3-phase rectifier for a 400-ampere output using six parallel cells in each arm of the 3-phase bridge. The average loading per cell is approximately 22 amperes. Since cells of much higher current rating are available, I assume that the number of cells in parallel was chosen to permit several failures without interruption of service. Is this set-up considered typical for normal application or was it chosen only to illustrate the principles of protective co-ordination?

The author mentions that current-limiting fuses for d-c bus interruption duty are available in ratings up to 1,200 amperes at 250 volts d-c. It may please him to learn that 5,000-ampere fuses are standard catalog items today. They are not of the same type as the fuses used for individual cell protection, but are particularly adaptable for co-ordinated d-c protection.

Reference

1. See pages 746–50 of this issue.

F. E. Gentry: I agree with Mr. Dortort that the forward characteristic of a semiconductor rectifier is not linear. For instance, silicon rectifiers have a current-versus-voltage characteristic which is exponential in nature. However, when the voltage drop across the cell exceeds about 0.8 volt the voltage drop can be represented to a very good approximation by a constant voltage drop plus a resistive drop. As is shown in Fig. 3, this representation is adequate for the surge times and current levels considered in the analysis.

If the semiconductor had a constant voltage drop at the junction which is independent of current, and if the bulk of the semiconductor were considered to have finite heat capacity and infinite thermal conductivity, and if it is attached to an insulating surface, then the solution would be a function of It. A more realistic model would include thermal conductivity. For this particular set of conditions an analysis has been made by Diebold (see reference 4 of my paper) where in his equation 20 it is shown that temperature is a function of heat input and $\sqrt{t}$. With a constant voltage drop across the junction and a square wave input of current, the heat generated would be IV, with the result that the temperature at the junction would be a function of $\sqrt{I^2t}$.

As pointed out in my paper, the surge capability of a rectifier will deviate from the I^2t constant relationship for times longer than l^2/α. This is expected since the geometry chosen for this analysis is not adequate. In a previous and much better analysis[1] the deviation from the constant I^2t relationship is predicted for times up to 10 milliseconds. As for the direction of the deviation, the constant I^2t rating will result in a conservative rating for times longer than the time from which the rating was determined, since the constant I^2t curve is somewhat steeper than the actual failure curve. This is evident from Fig. 5.

Reference

1. Surge Current Failure in Semiconductor Rectifiers, F. E. Gentry. *MEE Thesis*, Syracuse University, Syracuse, N. Y., June 1957.

F. W. Gutzwiller: The extensive oral discussion and the two formal discussions submitted on this paper and the paper by Mr. Gentry indicate the significant interest which exists on the subject of semiconductor rectifier protection. It is the author's hope that this vital matter will be the subject of additional papers and discussion.

The author is grateful to Mr. Lebens for his remarks. There is a certain philosophical humor in the fact that the fuse manufacturer has seen fit to convert the rectifier cell data into the terms and language being used by him for fuse purposes, and the author, who is affiliated with a rectifier cell manufacturer, has seen fit to convert the fuse data into terms more easily developed and understood by rectifier cell manufacturers.

Provided Mr. Lebens' method incorporates ample safety factor between fuse clearing and cell failure points, it appears that the two methods are compatible.

Comments on the first part of Mr. Dortort's discussion are made by Mr. Gentry in his companion paper. The following comments apply to the latter part of Mr. Dortort's discussion.

For noninterrupted service when a rectifier cell fails, the rectifier supply shown in Figs. 5 and 6 requires a minimum of four cells in parallel in each leg of the circuit. Three of the four cells are necessary to provide adequate fuse selectivity during a cell fault, as pointed out in the section of the paper "Fuse Co-ordination for Isolation of Defective Rectifier Cells." A minimum of one additional parallel cell in each leg is necessary as a "spare" element to permit satisfactory operation and fuse selectivity if one cell has already failed. In this particular example, two additional parallel cells in each leg were necessary to provide sufficient short-circuit capacity for the maximum d-c bus fault of 4,000 amperes. These two cells also provide ample safety margin between the characteristic of the breaker and the characteristics of the fuses and cells at lower level overloads. Mr. Dortort is correct in stating that a rectifier cell of slightly larger current rating may have reduced the number of parallel cells required in each leg. In any event, however, the minimum number of parallel cells is four.

It is very gratifying that the fuse industry is developing new types and ratings of current-limiting fuses to meet the requirements of semiconductor rectifiers. Mr. Dortort's very welcome notification of the availability of 5,000-ampere fuses for d-c protection is further evidence of this activity.

Spinning Disk Random Selectors

G. A. ROBERTS
ASSOCIATE MEMBER AIEE

THIS PAPER describes two spinning disk random selectors that generate equally-likely independent random numbers. One selector is a small, portable, single-decade unit, and the other is a 3-decade unit with the modulus adjustable from 10 to 1,000 in increments of 10.

Random numbers are used in numerous applications, some of which are, for example: selection of signals or input information for psychophysical experiments, simulation and testing of complex physical systems, setting of safe combinations, performance of Monte Carlo experiments, and all applications where random samples are needed. The random selectors described in this report are useful in all applications where a large quantity of independent random numbers are required or where random numbers other than those available in standard tables are desired.

Theory of Operation

The spinning disk random selector may be separated into two functional parts: an equally-likely number generator, and a random source, shown in the block diagram of Fig. 1. The equally-likely number generator consists of a recurrent pulse generator, an electronic counter of

---

Paper 58-534, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958. Manuscript submitted January 27, 1958; made available for printing May 21, 1958.

G. A. Roberts is with the University of Michigan, Ann Arbor, Mich.

This work was sponsored by the U. S. Army Signal Corps.

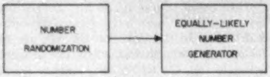
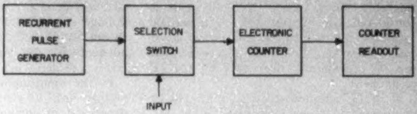


Fig. 1. Block diagram of spinning disk, equally-likely random number selector

Fig. 2 (right). Block diagram of equally-likely independent number generator



the proper modulus, and a number readout unit.

Assume the counter to have a modulus of N (i.e., N unique, stable states), and the period between pulses to be t. One modulus cycle then has a period T of Nt. The counter must pass through each of its N states for every modulus cycle. Because the input pulses are equally spaced, the counter will, on the average, be in each of its N states for the same period of time. If the counter is recurrently stopped in such a way that the stopping time is uniformly distributed over the modulus cycle, then the probability of a number being selected is equal to the faction of time the counter stays on that number.

If the selection time is random and the probability density of the selection time is uniformly distributed over T (the modulus period), then the numbers selected will be random and equally likely. If successive selection times are independent then successive numbers will be independent, with no patterning. A simplified block diagram of the equally-likely independent number generator is shown in Fig. 2.

The mechanism for randomizing the number selection must operate the selection switch at a random time and provide a uniform probability density function over the modulus cycle T. Assume that a device is available that delays an input pulse a random period of time. Further asume that the probability density function for this delay is approximated by a normal distribution. A sketch of this typical delay function is shown in Fig. 3.

The probability density function of Fig. 3 does not begin to approximate a uniform function. But this function is in the infinite time domain, whereas the desired function, the probability density function over the modulus period T, is over a finite time domain T. By taking the distribution of Fig. 3 and wrapping it around a cylinder of circumference T, and summing the ordinates, the probability density function for the interval T is obtained.* The larger the ratio of σ to T, the better is the approximation of the wrap-around probability density to a

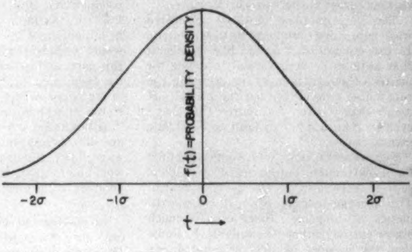


Fig. 3 (right). Normal delay function

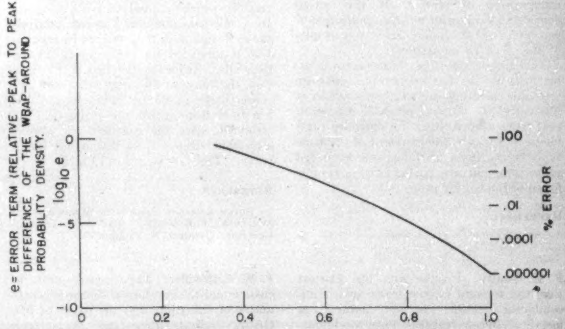


Fig. 4. Flatness of wrap-around probability density function based on normal delay function

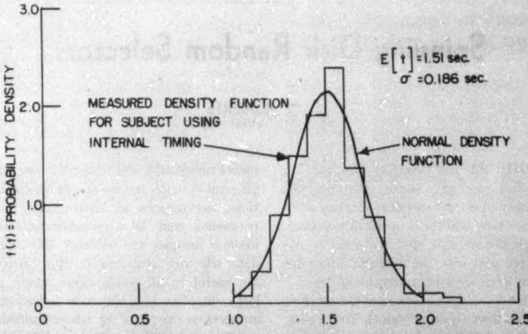


Fig. 5. Probability density curve for subject using internal timing

---

* T. G. Birdsall originated the wrap-around analysis.

Vol.
Pt. 1
1958

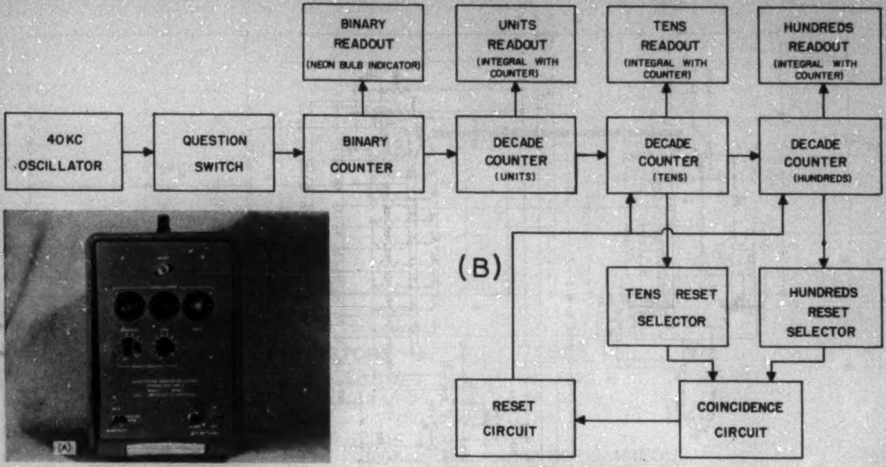

BINARY READOUT (NEON BULB INDICATOR)
UNITS READOUT (INTEGRAL WITH COUNTER)
TENS READOUT (INTEGRAL WITH COUNTER)
HUNDREDS READOUT (INTEGRAL WITH COUNTER)
40 KC OSCILLATOR
QUESTION SWITCH
BINARY COUNTER
DECADE COUNTER (UNITS)
DECADE COUNTER (TENS)
DECADE COUNTER (HUNDREDS)
(B)
TENS RESET SELECTOR
HUNDREDS RESET SELECTOR
RESET CIRCUIT
COINCIDENCE CIRCUIT
(A)

Fig. 6. Electronic spinning disk random selector, type C, model 1

A—Front view

B—Block diagram

uniform distribution. In Fig. 4 is shown a curve for the peak-to-peak difference of the probability density function over T as a function of the ratio of the standard deviation σ of the delay distribution curves to the modulus period T.

The random selectors described in this report are manually operated in the selection process. Therefore it is necessary to know if the foregoing assumptions are met when the randomizing device is a human operator. Two experiments have been conducted. The purpose of the first was to determine the smallest value of σ. A subject was given a stop watch and told to stop the watch as close to 4 seconds as possible. Several hundred trials were made with a result for a smaller σ of 35 milliseconds for the best of five subjects. In this experiment the subject has the best possible timing information available. A second experiment was run where the subject had no external timing information. This experiment utilized an electronic counter for time interval measurement. The subject was told to start the counter, count "one one-thousandth, two one-thousandths" and then stop the counter. The subject was to try to be as consistent in the formation of this time interval as possible. A value of 185 milliseconds was measured for σ. The results of this experiment are shown in Fig. 5 and, for comparison, a normal curve with the same mean and σ is plotted.

In summary, equally-likely random numbers are generated as follows: An electronic counter of modulus N is driven continuously by a recurrent pulse generator. Therefore, on the average, the counter is in each of its N possible states the same length of time and each number is equally likely. A selection switch is placed between pulse generator and counter so that the counter can be stopped when a selection is desired. Randomness is obtained by operating the counter at a relatively high rate with respect to the period between operations of the selection switch. Anything that operates the selection switch, or a human operator, will have some variance in the timing; if this variance is large with respect to the period for one cycle of modulus N, then independence of number selections is obtained.

Circuit Design

The two random selectors described are both designed for manual operation. The spinning disk random selector, type *C*, model 1, is a 3 decade, adjustable-modulus, semiportable unit. Fig. 6 shows a detailed block diagram of this unit which, as described here, was designed to meet the following requirements:

1. The selection process should be manual and occur no faster than once per second.
2. The maximum modulus of the counter should be 1,000 and should be adjustable from 10 to 1,000 in increments of 10.
3. The counter should use cold-cathode counter tubes because of their low cost, small size, low power requirements, and direct read-out.
4. In addition to the modulus N selection, an equally-likely binary output should be provided.

The use of cold-cathode counter tubes sets an upper limit to the operating speed of the counter. The type *GC-10-D* tube, which is a decade counter, has a maximum counting rate of 20 kc, and thus has an output of 2 kc. The *GC-10-D* does not have separate pin connections for each cathode; therefore it cannot be used as a preset counter. The *GS-10-C* has separate cathodes but is limited in counting speed to 4 kc. Therefore, the maximum counting speed can be achieved by using one *GC-10-D* and two *GS-10-C* tubes. This arrangement will also allow the desired modulus adjustment. (One-hundred-kc-per-second cold-cathode tubes are now available.)

For this particular counter, which has a maximum modulus of 1,000, the maximum modulus period is 0.02 second. If a criterion is selected of 1% flatness in the sequence 1% flatness in wrap-around probability density, then the smallest σ allowable is 0.03 second; see Fig. 4. Any σ larger than 0.03 second will result in a more uniform wrap-around probability density. The experimentally measured standard deviations of 0.035-second minimum, and 0.185 nominal, indicate that

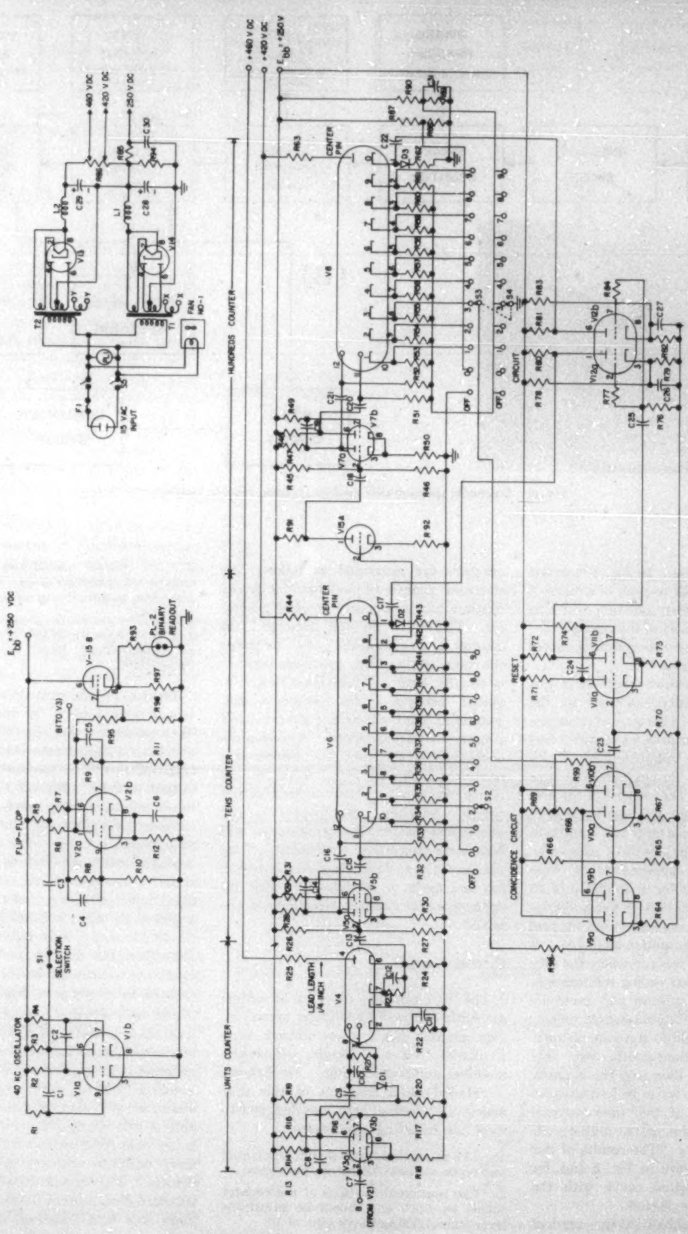
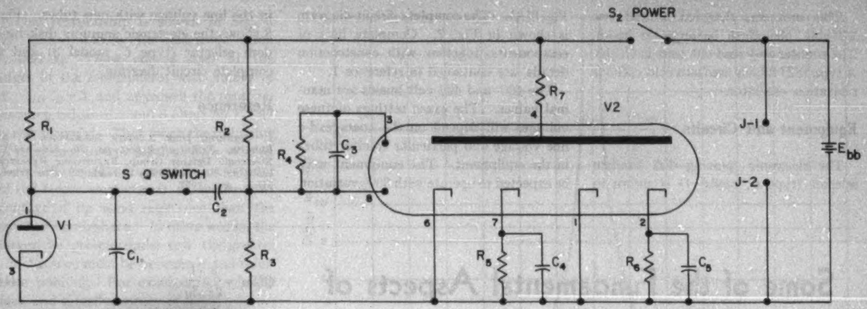

40 KC OSCILLATOR
V1a
V1b
S1
SELECTION SWITCH
FLIP-FLOP
V2a
V2b
B (TO V3)
V-15 B
PL-2 BINARY READOUT
E_{bb} +250 VDC
115 VAC INPUT
F1
PL-1
S5
FAN NO-1
T1
T2
V13
V14
L1
L2
480 VDC
420 VDC
250 VDC
UNITS COUNTER
TENS COUNTER
HUNDREDS COUNTER
B (FROM V2)
V3a
V3b
V4
LEAD LENGTH 1/4 INCH
V5a
V5b
V6
CENTER PIN
V7a
V7b
V8
V15A
S2
S3
S4
OFF
+480 VDC
+420 VDC
E_{bb} +250 V
COINCIDENCE CIRCUIT
V9a
V9b
V10a
V10b
RESET
V11a
V11b
V12a
V12b

Fig. 7. Circuit diagram for electronic random selector spinning disk, type C, model 1



RESISTORS

R1	3.3 MEG	1/2 W
R2	12 MEG	1/2 W
R3	2.2 MEG	1/2 W
R4,5,6	220K	1/2 W
R7	330K	1/2 W

CONDENSERS

C1,2	500 PFD	600V
C3,4,5	100 PFD	600V

TUBES

V1	5823
V2	GC-10-D

MISCELLANEOUS

S-1 SWITCHCRAFT 3033L (MODIFIED FOR MOMENTARY AND HOLD OPERATION)

S-2 BAT HANDLE SPST GC-1330

E_{bb} 2-240 V STROBOFLASH BATTERIES (IN SERIES)

Fig. 8. Electronic spinning disk random selector, type C, model 3

the manual operation will certainly meet this criterion.

If a "safety factor" is desired, the operator of the equipment should not watch the counter tubes prior to actuating the selection switch, since timing information may be obtained from the hundreds counter. The result of receiving this information may be a reduction in σ, although undoubtedly it will still be above the minimum allowable of 30 milliseconds.

The modulus of the counter is set by two switches. The switches select one cathode on the tens counter and one on the hundreds counter. The outputs from the switches are applied to a coincidence circuit in such a way that when a voltage exists between both the selected cathodes and ground, the counter is forced to reset. Thus, any modulus between 10 and 1,000 in increments of 10 may be obtained.

The output of the binary unit is applied to the modulus N counter. A *GS-10-D* counter is used for the units decade. (This tube may be obtained from the Atomic Instrument Company of Cambridge, Mass.) All cathodes except the output cathode are tied together in this tube; therefore, only decade counting can be obtained. The output indicator is an integral part of the counter, and it utilizes a gaseous discharge glow that can appear in one of ten different positions on a circular path. The angular position indicates the count. The output cathode and resistor produce one output pulse for every ten input pulses to the counter. This pulse is coupled to the input of the next counter. The physical orientation of the tube relative to the bezel should be such that the output cathode lines up with the O on the bezel.

The tens and hundreds counters are identical to each other and are similar to the units counter. The primary difference is that the tens and hundreds counters have the lower counting speed of 4 kc, and all cathodes are brought out separately. A cathode is illuminated by a current flowing through it, and a voltage drop is developed across the cathode resistor. Since all cathodes are brought out separately, and each has a separate resistor, the location of the cathode glow can be identified by the presence of a voltage. This provides a convenient method for identifying the state of the counter. In this particular application an 11-position switch is used with the tens and hundreds counters to select the cathodes that trigger the automatic reset circuit. As mentioned, when coincidence of these two switch outputs occurs, the tens and hundreds counters are reset to "000." The coincidence circuit drives a reset circuit in order to provide a sufficiently long reset pulse. The reset circuit is a multivibrator followed by a reset amplifier for the tens counter and one for the hundreds counter. Since the reset circuit insures that resetting goes to completion, there is no bias towards the last number in the modulus, or "000."

The spinning disk random selector, type *C*, model 3, was designed as a lightweight, handy unit and is used for small applications such as setting safe combinations or where a few selections are needed for a particular problem. The input frequency to the counter is approximately 500 cycles per second; this results in a modulus period of approximately 20 milliseconds. Therefore, better than 1% peak-to-peak flatness will be achieved in the wrap-around probability density function.

This unit was designed around two 240-volt photoflash batteries in series. The counter and read-out tube is *GC-10-A* type *5823* tube is used as a cold-cathode relaxation oscillator.

Equipment and Circuits

The electronic spinning disk random selector (type *C*, model 1) is shown in Fig. 6(A). The complete circuit diagram is shown in Fig. 7. Complete lists of components, together with construction details, are contained in reference 1.

The 460- and 420-volt busses are nominal values. The exact settings of these voltages will depend on the average d-c line voltage and particular counter tubes in the equipment. The equipment may be expected to operate with 30% variation in the line voltage with new tubes. Fig. 8 shows the electronic spinning disk random selector (type *C*, model 3) and a complete circuit diagram.

Reference

1. Spinning Disk Random Selectors, G. A. Roberts. *Technical Report no. 39, 2262-151-T*, Electronic Defense Group, Engineering Research Institute, The University of Michigan, Ann Arbor, Mich., Apr. 1957.

Some of the Fundamental Aspects of Ground Resistance Measurements

E. B. CURDTS
MEMBER AIEE

THE PRACTICAL measurement of the resistance to earth of a very small symmetrical electrode is a fairly simple procedure and does not call for a great deal of knowledge of the fundamentals involved. However, as such electrodes increase in size and become more complex in their shape, there is greater need for a better understanding of the basic principles.

It is the purpose of this paper to present some of these basic principles in their simplified form to assist in making realistic measurements. To accomplish this objective a homogeneous single-layer earth will be assumed throughout this discussion. These simplifications will not affect the basic principles involved, but on the other hand there can be no theoretical substitute for actual field measurements. However, a better understanding of the problem should greatly assist in anticipating the measurement procedures.

Statement of the Measurement Problem

When a resistance measurement is to be made on a ground grid of a certain size and shape, the following questions immediately arise:

1. At what distance from the ground under test should the current reference electrode be placed?
2. Where should the potential reference electrode be placed in relation to the current reference electrode and the ground under test?
3. What effect does the size and shape of the ground under test have on the answers to questions 1 and 2?

Basic Analysis

The network is shown schematically in Fig. 1. As a first analytical step in obtaining a theoretical answer to these three questions the mutual resistance to true earth can be calculated for a sink (point 1) and a current source (point 3) spaced at a remote or infinite distance from point 1, as given by

$$Q(S_{12})=\frac{\rho}{2\pi}\frac{1}{S_{12}} \qquad (1)$$

where S_{12} is the spacing between point 1 and a movable point 2. This mutual resistance is shown plotted in Fig. 2 for $\rho/2\pi=1{,}000$, ρ being the resistivity of the earth. It approaches infinity near point 1 and approaches zero near the remote current source 3.

In Fig. 3 hemispherical metal electrodes with radii r_1 and r_2 are shown around point 1, together with the same mutual resistance curve as given in Fig. 2. The resistance to true earth is then determined by a point on this curve representing, in terms of S_{12}, the radius of the hemisphere chosen. Hemisphere radius $r_1=200$ feet results in a total resistance of 5 ohms from its edge to true earth, and $r_2=500$ feet results in a total resistance of 2 ohms.

The mathematical expression for the relationship of the hemisphere under test and its incremental resistance at any distance S_{12} is then given by

$$R_{(1-3)(1-2)}=\frac{\rho}{2\pi}\left(\frac{1}{S_r}-\frac{1}{S_{12}}\right) \quad S_{12}\overline{>}S_r \qquad (2)$$

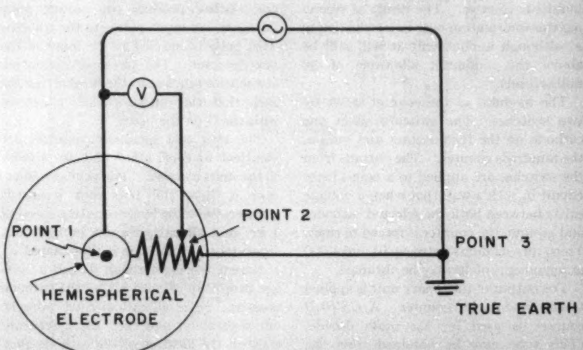


Fig. 1. Schematic network for measurement of resistance to true earth of hemispherical electrode

---

Paper 58-106, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted November 6, 1957; made available for printing May 20, 1958.

E. B. Curdts is with the James G. Biddle Company, Philadelphia, Pa.

This relationship is shown in Fig. 3 by heavy curves for electrode radii r_1 and r_2. These curves start at the edges of the hemispheres with values of $R_{(1-3)(1-2)}=0$, and approach the total resistance values of 2 and 5 ohms for r_1 and r_2 respectively. They reach the full value of resistance to true earth at $S_{12}=\infty$. At $S_{12}=1{,}500$ feet the incremental resistance of the larger electrode is a smaller percentage of its total resistance than the smaller hemisphere. In other words, the larger the ground under test, the greater the distance must be between it and reference point 2. For example, for $r_1=200$ feet and a test accuracy of 90%,

$$R_{(1-3)(1-2)}=\frac{1{,}000}{200}\times 0.90=4.5$$

$$=\frac{1{,}000}{200}-\frac{1{,}000}{S_{12}}$$

$$S_{12}=\frac{1{,}000}{0.5}=2{,}000 \text{ feet}$$

For $r_2=500$ feet and the same accuracy

$$R_{(1-3)(1-2)}=\frac{1{,}000}{500}\times 0.90=1.8$$

$$=\frac{1{,}000}{500}-\frac{1{,}000}{S_{12}}$$

$$S_{12}=\frac{1{,}000}{0.2}=5{,}000 \text{ feet}$$

Stated differently, the percentage of the total resistance obtained for any distance S_{12} is given by

$$\% R_\infty=\frac{\frac{\rho}{2\pi}\left(\frac{1}{S_r}-\frac{1}{S_{12}}\right)}{\rho/2\pi\ 1/S_r}100$$

$$=\left(1-\frac{S_r}{S_{12}}\right)100 \tag{3}$$

Inspection of equation 3 shows that any error resulting from any one location of point 2 is a function of S_r only, and is not affected by earth resistivity.

Summation of Mutual Resistances

It is to be emphasized that the foregoing analysis is based on the current source 3 being spaced an infinite distance from the ground under test. Considering the problem for the case where source 3 is located at a more practical distance from point 1, the mutual resistances $Q(S_{13})$ and $Q(S_{23})$ have a definite influence on the measurement, the complete expression being given by

$$R_{(1-3)(1-2)}=\frac{\rho}{2\pi}\left(\frac{1}{S_r}-\frac{1}{S_{12}}-\frac{1}{S_{13}-S_r}+\frac{1}{S_{23}}\right) \tag{4}$$

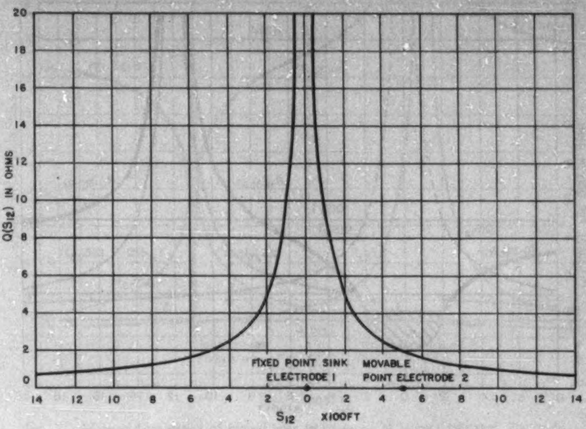


Fig. 2. Mutual resistance to true earth of point electrode

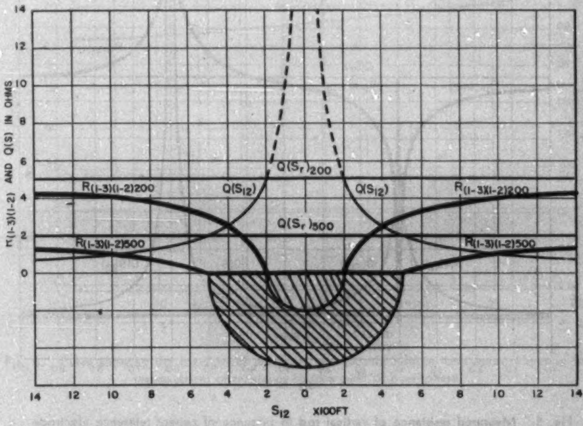


Fig. 3. Mutual resistance to true earth of hemispheres having radii of 200 and 500 feet

The mutual resistance between points 1 and 3 is a single value for a fixed location of point 3, and is given by

$$Q(S_{13})=\frac{\rho}{2\pi}\ \frac{1}{S_{13}-S_r} \tag{5}$$

The mutual resistance between the movable point 2 and the fixed point 3 is a variable given by

$$Q(S_{23})=\frac{\rho}{2\pi}\ \frac{1}{S_{23}} \tag{6}$$

Fig. 4 shows an electrode at 1 with $r_1=200$ feet, and a current point electrode 3 at a distance of 1,200 feet from point 1. By equation 5

$$Q(S_{13})=1{,}000\left(\frac{1}{1{,}200-200}\right)=1.0 \text{ ohm}$$

Since by equation 4 this value is subtracted in determining $R_{(1-3)(1-2)}$, it is readily seen that the proximity of electrode 3 to point 1 can adversely affect

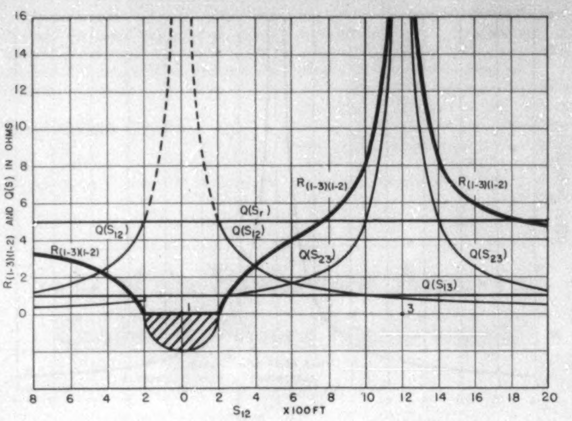


Fig. 4. Measured resistance of hemisphere in presence of current reference electrode

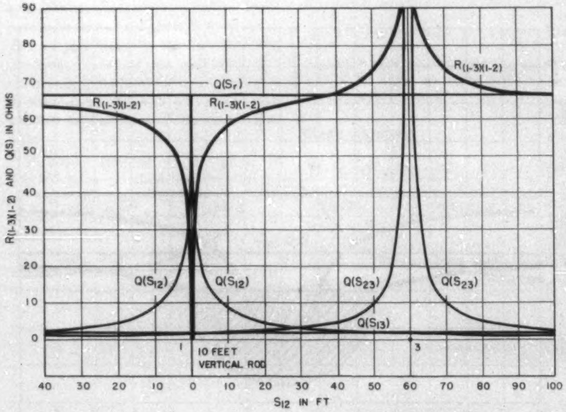


Fig. 5. Measured resistance of vertical rod in presence of current reference electrode

the measurement. The constant value of $Q(S_{13}) = 1.0$ ohms is shown in Fig. 4.

Fig. 4 also shows the variable mutual resistance $Q(S_{23})$ as calculated from equation 6, together with $Q(S_{12})$ from Fig. 2.

The heavy curves show the resistance values $R_{(1-3)(1-2)}$ as calculated from equation 4 and based on the potential reference point 2 being moved along a straight line from the far side of point 1 to the far side of point 3. A study of these curves reveals interesting information concerning the values of $R_{(1-3)(1-2)}$ for three of the many possible locations of point 2 in relation to the current source and sink. For example, if S_{12} is chosen as 50 and 150% of S_{13}, 600 and 1,800 feet from point 1 in Fig. 4, the following values are obtained:

Position of Electrode 2	No. of Feet From Point 1	Ohms
(A) On far side of point 1	600	2.88
(B) Between points 1 and 3	600	4.0
(C) On far side of point 3	1,800	5.11

In the case of position (A), $R_{(1-3)(1-2)}$ will be less than the resistance to true earth by at least $Q(S_{13}) = 1.0$ ohm even after increasing S_{13} to infinity. In the cases of positions (B) and (C), $R_{(1-3)(1-2)}$ can be made to equal the resistance to true earth at some value of S_{12}, to be determined.

Resistance to Earth Versus Measured Resistance

If in equation 4 $Q(S_{12})+Q(S_{13})$ can be made to equal $Q(S_{23})$, they will cancel and $R_{(1-3)(1-2)}$ will equal $Q(S_r)$, the true ground resistance. It can be shown mathematically that this is true, in the simplified case being analyzed here, for the (B) and (C) locations of electrode 2 as follows:

$$\left.\begin{array}{ll}\text{(B).}\ S_{12}=\dfrac{-1+\sqrt{5}}{2}S_{13}=0.62\,S_{13} & (7)\\ \text{(C).}\ S_{12}=\dfrac{1+\sqrt{5}}{2}S_{13}=1.62\,S_{13} & (8)\end{array}\right\} S_r \ll S_{13}$$

Equations 7 and 8 show that if the potential reference electrode is placed out from point 1 along the straight line through point 3 at either 62% or 162% of the spacing between points 1 and 3, then a resistance will be measured which is equal to the total resistance of the earth surrounding the ground under test. Such a simple mathematical relationship would apply precisely only in the case of point electrodes throughout. It also applies with negligible error to hemispherical or other shaped grounds if they are not too large in relation to S_{13}. For example, a single vertically driven rod or a buried wire (two commonly used electrode arrangements) will be similar to the hemisphere in this respect, their length influencing this relationship even less than the radius of the hemisphere.

Curves for a single vertically driven rod are shown in Fig. 5. It will be seen that the $Q(S_{13})$ curve crosses the $Q(S_{23})$ curve approximately midway between points 1 and 3. This means that in the case of a single driven rod, at least, the total resistance of the earth surrounding it should be found at approximately 62 and 162% of the straight-line distance through points 1 and 3 even for rod lengths which are appreciable in relation to S_{13}.

The case of a single straight buried wire, measured as shown in Fig. 6, is plotted in Fig. 7. It will be seen that the $Q(S_{12})$ curve also crosses the $Q(S_{23})$ curve approximately midway between points 1 and 3. As in the case of the vertical rod, this means that the total resistance of the earth surrounding a

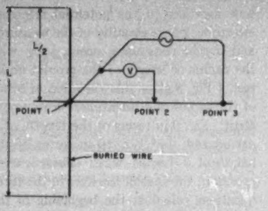


Fig. 6. Schematic network for measurement of resistance of buried wire in presence of current reference electrode

buried wire, when measured as shown in Fig. 6, should be found at approximately 62% and 162% of the straight-line distance through points 1 and 3 even for wire lengths which are appreciable in relation to S_{13}.

Various authors in writing on this subject have stated that the total resistance of the ground under test can be determined by placing point 2 on the opposite side from point 3. Figs. 4, 5, and 7 show that such measurements will always be less than the true resistance by the value of $Q(S_{13})$ and the difference between the values of $Q(S_{12})$ and $Q(S_{23})$.

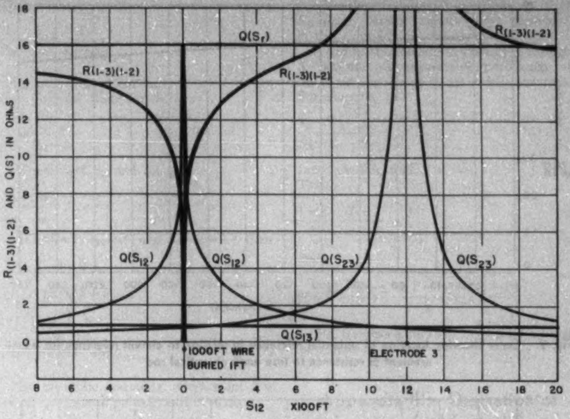


Fig. 7. Measured resistance of buried wire in presence of current reference electrode

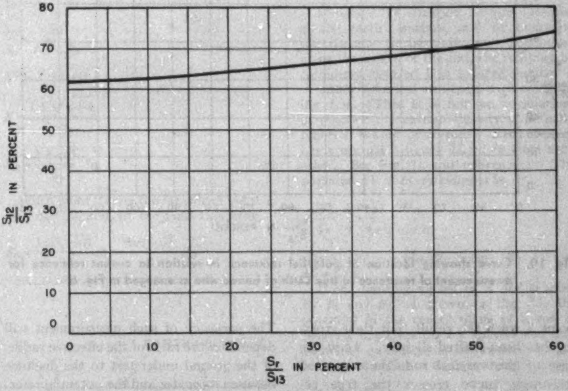


Fig. 8. Curve showing the location of potential reference in relation to current reference for measurement of resistance to true earth of hemisphere

Summary

Several examples of how ground resistance measurements are affected by the size and shape of the ground under test, and the location of the reference electrodes, are summarized as follows:

Fig. 2 shows the basic characteristic curve of the resistance to true earth at increasing distances from a point designated by electrode 1. It indicates that most of the resistance exists within a very small distance from the point. The curve, under conditions of constant current, also represents the voltage gradient for any radial increment of earth surrounding point 1, indicating much higher gradients nearer the ground under test.

Fig. 3 shows how an increase in ground grid radius decreases the total resistance to earth and the slope of the resistance-distance curve. The figure also reveals that ground electrodes of larger radii require a greater distance to reach their final resistance values.

Fig. 4 shows how the shape of the measured resistance curve is influenced by the various mutual resistances involved as the current reference electrode is brought from infinity to the vicinity of the ground under test, as shown in Fig. 1. It also indicates the results to be expected by placing the potential reference electrode at various locations along a straight line from the far side of ground electrode 1 to the far side of current reference electrode 3. At only two locations of potential reference electrode 2 along this line does the measured resistance represent the total resistance of ground electrode 1, at approximately 62% and 162% of the space between points 1 and 3.

Fig. 5 shows the results of replacing the hemispherically shaped electrode with a rod driven vertically into the earth. Here the mutual resistance of the rod as a function of distance is somewhat different from that of point 3, with the result that their crossing has been shifted slightly. This causes the measured resistance curve to cross the true resistance line at a distance somewhat different from that given for a hemispherical electrode.

Fig. 7 shows the results of replacing the hemispherically shaped electrode with a buried wire measured as shown in Fig. 6. Here the mutual resistance of the buried wire as a function of distance is also somewhat different from that of

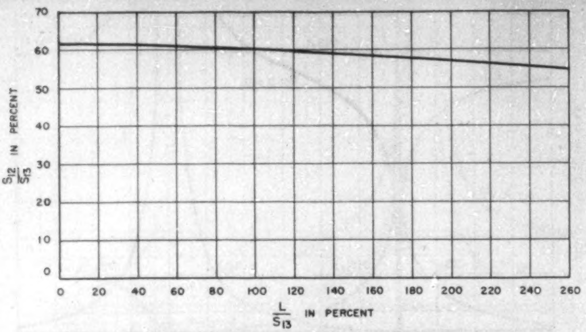


Fig. 9. Curve showing location of potential reference in relation to current reference for measurement of resistance to true earth of vertical rod

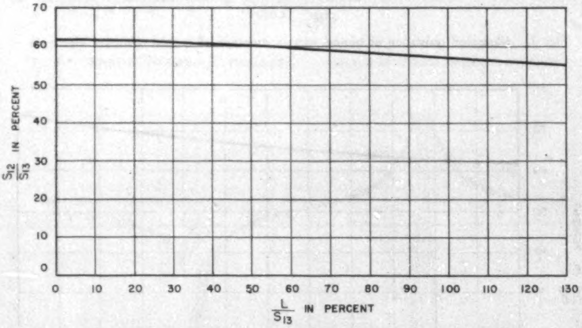


Fig. 10. Curve showing location of potential reference in relation to current reference for measurement of resistance to true earth of buried wire as arranged in Fig. 6

point 3, with the result that their crossing has been shifted slightly. As in the case of the vertical rod, the measured resistance curve crosses the true resistance line at a distance somewhat different from that given for a hemispherical electrode.

Practical Conclusions

The conclusions to be drawn from this analysis which theoretically answer the questions raised at the beginning of this paper are as follows:

If the current reference electrode is located at a distance so remote from the ground under test that its mutual resistances can be neglected, the measured resistance for any given size and shape of ground under test approaches the actual resistance to true earth as the distance of the potential reference electrode in relation to the ground under test is increased. The accuracy of such measurement will depend on the ratio of the effective radius of the ground under test to the distance between its center and the potential reference electrode.

Because so remote a location of the current reference electrode would ordinarily be impractical, the influence of its mutual resistances must be considered. Fortunately, these mutual resistances become additive in their effect on the measured resistance so that, at a point in the order of 60% of the distance between the effective center of the ground under test and the current reference electrode, the measured and actual resistances become equal. Although the theoretical analysis shows that the measurement can also be made on the far side of the current reference electrode, there would appear to be no practical justification for using a test connection of this extent.

Curves given in Figs. 8, 9, and 10 show how locations of the potential reference electrode, where equality of the measured and actual resistances occur, vary with the radius or length of the ground under test. Fig. 8 gives this variation in terms of the radius of a hemisphere or its equivalent; Fig. 9 in terms of the length of a driven rod, and Fig. 10 in terms of the length of a buried wire. These curves represent the specific answers to the three questions raised at the beginning of the paper, and provide the theoretical solution to the measurement problem.

Appendix I. Derivation of Equations 1, 2, and 4

The derivation of equation 1 for point electrodes 1 and 2 with point 3 at infinity follows: The radial current density at distance S from point 1 is given by

$$J=I/2\pi S^2$$

and the electric intensity at the distance S for an earth resistivity of ρ, will be

$$E(S)=I\rho/2\pi S^2$$

The potential at the distance S is then

$$\int_{S_{12}}^{\infty} E(S)dS=I\rho/2\pi S_{12}$$

and the mutual resistance of points 1 and 2 is

$$Q(S_{12})=\frac{\rho}{2\pi}\,\frac{1}{S_{12}} \tag{1}$$

The derivation of equation 2 for a hemispherical electrode surrounding point 1 follows: The voltage between points 1 and 2 due to the presence of a metallic hemispherical electrode at point 1 is

$$\int_0^{S_r} \rho_0 I/2\pi S^2 dS+\int_{S_r}^{S_{12}} \rho I/2\pi S^2 dS$$

where S_r is the radius of the hemisphere and ρ_0 is its resistivity. The measured resistance then becomes

$$R_{(1-3)(1-2)}=Q(S)_1=\int_{S_r}^{S_{12}} \rho/2\pi S^2 dS$$
$$=\frac{\rho}{2\pi}\left(\frac{1}{S_r}-\frac{1}{S_{12}}\right) \quad \rho_0\ll\rho \tag{2}$$

The derivation of equation 4 for a hemispherical ground and with point 3 moved toward point 1 follows: The presence of point 3 in the vicinity of point 1 introduces an additional voltage between points 1 and 2 which is given by

$$-\int_{S_{13}-S_r}^{S_{23}} \rho I/2\pi S^2\, dS-\int_{S_{13}}^{S_{13}-S_r} \rho_0 I/2\pi S^2 dS$$
$$=\int_{S_{23}}^{S_{13}-S_r} \rho I/2\pi S^2 dS$$
$$=\frac{\rho I}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{S_{13}-S_r}\right) \quad \rho_0\ll\rho$$

Therefore

$$Q(S)_3=\frac{\rho}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{S_{13}-S_r}\right) \text{ and}$$

$$R_{(1-3)(1-2)}=Q(S)_1+Q(S)_3$$

$$=\frac{\rho}{2\pi}\left(\frac{1}{S_r}-\frac{1}{S_{12}}-\frac{1}{S_{13}-S_r}+\frac{1}{S_{23}}\right) \qquad (4)$$

where $R_{(1-3)(1-2)}$ equals the measured resistance, the current passing through the earth between points 1 and 3, and the voltage measured between points 1 and 2.

Appendix II. Derivation of Equations 7 and 8

For $R_{(1-3)(1-2)}$ to equal $Q(S_r)$ the remaining mutual resistance terms in equation 4 must equal zero. Therefore

$$\frac{1}{S_{23}}=\frac{1}{S_{12}}+\frac{1}{S_{13}-S_r}$$

When point 2 is between points 1 and 3, and since points 1, 2, and 3 are in a straight line,

$$S_{12}=S_r+S_{r2}$$

$$S_{13}-S_r=S_{r2}+S_{23}$$

$$S_{13}=S_{12}+S_{23}$$

Therefore

$$\frac{1}{S_{23}}=\frac{1}{S_{12}}+\frac{1}{S_{r2}+S_{23}}$$

$$\frac{S_{23}}{S_{12}}=\frac{S_{r2}}{S_{r2}+S_{23}}=\frac{S_{12}-S_r}{S_{12}-S_r+S_{23}}$$

$$=\frac{S_{12}-S_r}{S_{13}-S_r}$$

Where $S_r \ll S_{13}$ and S_{12}:

$$\frac{S_{23}}{S_{12}}=\frac{S_{12}}{S_{13}}$$

Let $S_{13}=1$, $S_{12}=X$, and $S_{23}=1-X$; then

$$X^2+X-1=0$$

$$X=\frac{-1\pm\sqrt{1+4}}{2}$$

$$S_{12}=\frac{-1+\sqrt{5}}{2}\,S_{13} \qquad (7)$$

By the same reasoning, when point 2 is on the far side of point 3, and $S_r \ll S_{13}$

$$S_{12}=S_{13}+S_{23}$$

$$\frac{1}{S_{12}-S_{13}}=\frac{1}{S_{12}}+\frac{1}{S_{13}}$$

$$\frac{S_{12}}{S_{13}}=\frac{S_{13}}{S_{12}-S_{13}}$$

$$X^2-X-1=0$$

$$X=\frac{1\pm\sqrt{1+4}}{2}$$

$$S_{12}=\frac{1+\sqrt{5}}{2}\,S_{13} \qquad (8)$$

When point 2 is on the far side of point 1, $S_{23}=S_{12}+S_{13}$ and, by the same reasoning,

$$S_{12}=\frac{1\pm\sqrt{-3}}{2}\,S_{13}=\frac{1\pm j\sqrt{3}}{2}\,S_{13}$$

an imaginary quantity, indicating that $R_{(1-3)(1-2)}$ cannot equal $Q(S_r)$ anywhere in this direction on the earth's surface.

It can further be shown that even though S_r becomes an appreciable percentage of S_{13} equations 7 and 8 are fairly accurate. For example, in

$$\frac{S_{23}}{S_{12}}=\frac{S_{12}-S_r}{S_{13}-S_r}$$

Let $S_{12}=XS_{13}$ and $S_r=YS_{13}$.
Then

$$\frac{1-X}{X}=\frac{X-Y}{1-Y}$$

and

$$X=\frac{2Y-1+\sqrt{(2Y-1)^2+4(1-Y)}}{2}$$

Fig. 8 shows a curve of S_{12}/S_{13} for a range of values of S_r/S_{13}.

Appendix III. Derivation of Equation for Vertical Rod Electrode

To derive the equation for the potential at the earth's surface, and by assuming the current to be uniformly distributed along the length of the rod, the rod may be considered divided into a large number n of nearly spherical elements, each conducting I/n. (This is a fair approximation; in practice the current density is actually higher at the end of the rod.) Each element has a mutual distance $dL=L/n$ from each other. See Fig. 11 and reference 1. The potential dV of every element is

$$\rho\,\frac{I/n}{4\pi}\,\frac{1}{y}=\frac{\rho I}{4\pi L}\,\frac{d\alpha}{\sin\alpha}$$

where $\sin\alpha=y\,d\alpha/dL$ from Fig. 11 and $dL=L/n$.

If the limiting values of α are denoted by β_1 and β_2, as shown in Fig. 11, the potential in the center plane of a rod of length 2 L, containing 2 n spheres and thus including its image, is determined by the length and radius of the rod and given by

$$\frac{\rho I}{4\pi L}\int_{\beta_1}^{\pi-\beta_1}\frac{d\alpha}{\sin\alpha}=\frac{\rho I}{2\pi L}\ln\left(\cot\frac{\beta}{2}\right)$$

$$=\frac{\rho I}{2\pi L}\left(\ln\frac{2L}{a}\right)$$

where $L \gg a$

Thus the voltage between points 1 and 2 with a current reference at point 3 is

$$V=\frac{\rho I}{2\pi L}\left(\ln\frac{2L}{a}\right)-\frac{\rho I}{4\pi L}\int_{\beta_2}^{\pi-\beta_2}\frac{d\alpha}{\sin\alpha}+\frac{\rho I}{2\pi}\int_{S_{23}}^{S_{13}-a}\frac{dX}{X^2}$$

$$=\frac{\rho I}{2\pi L}\left(\ln\frac{2L}{a}-\ln\frac{\sqrt{L^2+S_{12}^2}+L}{S_{12}}\right)+\frac{\rho I}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{S_{13}-a}\right)$$

and

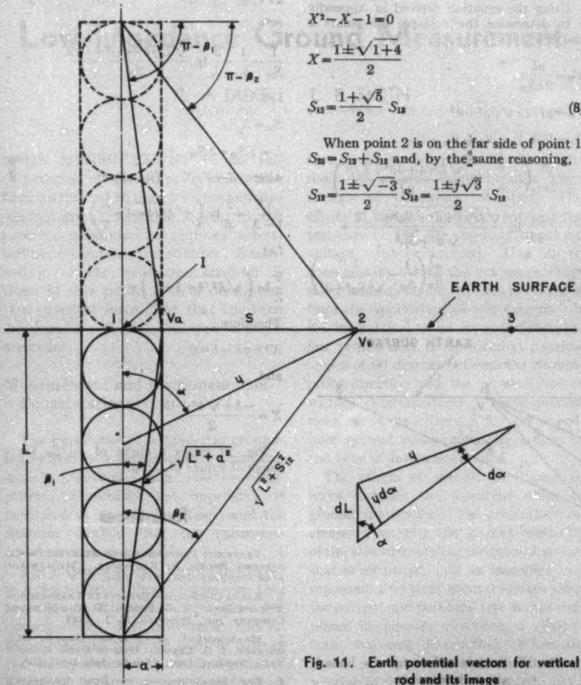


Fig. 11. Earth potential vectors for vertical rod and its image

$$R_{(1-3)(1-2)}=\frac{\rho}{2\pi}\left(\frac{1}{L}\ln\frac{2L}{a}-\right.$$

$$\left.\frac{1}{L}\ln\frac{\sqrt{L^2+S_{12}^2}+L}{S_{12}}+\frac{1}{S_{23}}-\frac{1}{S_{13}-a}\right)$$

To show that equations 7 and 8 are fairly accurate for various lengths of driven rods as well as for hemispheres, let

$$\frac{\rho}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{S_{13}-a}-\frac{1}{L}\ln\frac{\sqrt{L^2+S_{12}^2}+L}{S_{12}}\right)=0$$

or

$$\frac{1}{S_{23}}-\frac{1}{S_{13}}=\frac{1}{L}\ln\frac{\sqrt{L^2+S_{12}^2}+L}{S_{12}}$$

where $S_{13}\gg a$
Let

$$S_{12}=XS_{13}$$

$$L=YS_{13}=ZS_{12}$$

where $Z=Y/X$. Therefore

$$\frac{X^2}{1-X}=\frac{1}{Z}\ln(\sqrt{1+Z^2}+Z)$$

Let

$$\frac{1}{Z}\ln(\sqrt{1+Z^2}+Z)=b$$

Therefore

$$X^2+bX-b=0$$

and

$$X=\frac{-b+\sqrt{b^2+4b}}{2}$$

Fig. 9 shows a curve of S_{12}/S_{13} for a range of values of L/S_{13}.

Appendix IV. Derivation of Equation for Buried Horizontal Wire Electrode

Following the same method of solution of the problem as given in Appendixes I and III, and assuming the connection to the buried wire to be insulated, the geometry of the wire and its image is determined as shown in Fig. 12. Here the limiting values of the angles involved are determined by the length, depth, and radius of the wire, which determine the voltages involved, and the measured resistance at point 2 in the center plane or earth surface with a current reference at point 3.

Using the equation derived in Appendix III to determine the voltages of the wire, its image, and that of point 2

$$V_1=\frac{\rho I}{2\pi L}\ln\left(\cot\frac{\beta_1}{2}\right)$$

$$V_i=\frac{\rho I}{2\pi L}\ln\left(\cot\frac{\beta_i}{2}\right)$$

$$V_{12}=\frac{\rho I}{\pi L}\ln\left(\cot\frac{\beta_2}{2}\right)$$

Using the equation derived in Appendix I to determine the voltages for points 2 and 3

$$V_{23}=\frac{\rho I}{2\pi S_{23}}$$

$$V_{13}=\rho I/2\pi\sqrt{S_{13}^2+d^2}$$

$$V_s=V_1+V_i-V_{12}+V_{23}-V_{13}$$

$$=\frac{\rho I}{2\pi L}\left(\ln\frac{L}{a}+\ln\frac{\sqrt{L^2/4+4d^2}+L/2}{2d}\right)-$$

$$\frac{\rho I}{\pi L}\ln\frac{\sqrt{L^2/4+d^2+S_{12}^2}+L/2}{\sqrt{d^2+S_{12}^2}}+$$

$$\frac{\rho I}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{\sqrt{S_{13}^2+d^2}}\right)$$

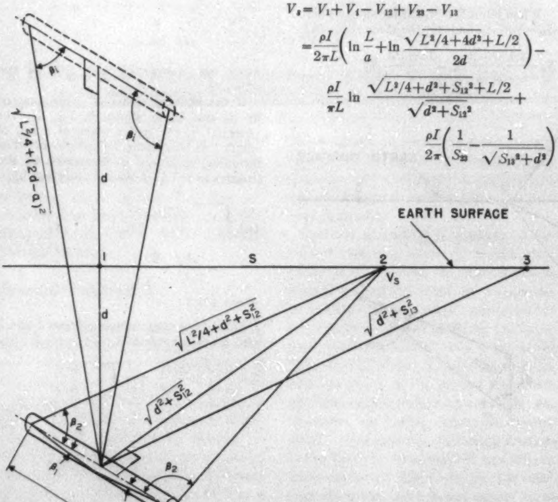


Fig. 12. Earth potential vectors for buried wire and its image

where $a\ll L$, and $a\ll d$

$$R_{(1-3)(1-2)}=\frac{\rho}{2\pi}\times$$

$$\left(\frac{1}{L}\ln\frac{L\sqrt{L^2/4+4d^2}+L^2/2}{2ad}-\right.$$

$$\frac{2}{L}\ln\frac{\sqrt{L^2/4+d^2+S_{12}^2}+L/2}{\sqrt{d^2+S_{12}^2}}-$$

$$\left.\frac{1}{\sqrt{S_{13}^2+d^2}}+\frac{1}{S_{23}}\right)$$

$$=\frac{\rho}{2\pi}\left(\frac{1}{L}\ln\frac{L^2}{2ad}-\frac{2}{L}\ln\frac{\sqrt{L^2/4+S_{12}^2}+L/2}{\sqrt{d^2+S_{12}^2}}-\right.$$

$$\left.\frac{1}{\sqrt{S_{13}^2+d^2}}+\frac{1}{S_{23}}\right)$$

where $d\ll L$

$$=\frac{\rho}{2\pi}\left(\frac{1}{L}\ln\frac{L^2}{2ad}-\frac{2}{L}\times\right.$$

$$\left.\ln\frac{\sqrt{L^2/4+S_{12}^2}+L/2}{S_{12}}-\frac{1}{S_{13}}+\frac{1}{S_{23}}\right)$$

where $d\ll S_{12}$

To show that equations 7 and 8 are fairly accurate for various lengths of buried wire, let

$$\frac{\rho}{2\pi}\left(\frac{1}{S_{23}}-\frac{1}{S_{13}}-\frac{2}{L}\ln\frac{\sqrt{L^2/4+S_{12}^2}+L/2}{S_{12}}\right)=0$$

or

$$\frac{1}{S_{23}}-\frac{1}{S_{13}}=\frac{2}{L}\ln\frac{\sqrt{L^2/4+S_{12}^2}+L/2}{S_{12}}$$

Let

$$S_{12}=XS_{13}$$

$$L=YS_{13}=ZS_{12}$$

where $Z=Y/X$. Therefore

$$\frac{X^2}{1-X}=\frac{2}{Z}\ln\left(\sqrt{Z^2/4+1}+\frac{Z}{2}\right)$$

Let

$$\frac{2}{Z}\ln\left(\sqrt{Z^2/4+1}+\frac{Z}{2}\right)=b$$

Therefore

$$X^2+bX-b=0$$

and

$$X=\frac{-b+\sqrt{b^2+4b}}{2}$$

Fig. 10 shows a curve of S_{12}/S_{13} for a range of values of L/S_{13}.

References

1. TRANSIENT PERFORMANCE OF ELECTRIC POWER SYSTEMS (book), R. Rüdenberg. McGraw-Hill Book Company, Inc., New York, N. Y.

2. EARTH CONDUCTION EFFECTS IN TRANSMISSION SYSTEMS (book), E. D. Sunde. D. Van Nostrand Company, Inc., Princeton, N. J., 1948.

3. MEASUREMENT OF STATION GROUND RESISTANCE, J. G. Cassan. *Ontario Hydro Research News*, Toronto, Ont., Canada, July-Dec. 1956.

4. THE MEASUREMENT OF LOW RESISTANCE

Earth Connections, G. F. Tagg. *Electrical Times*, London, England, Apr. 1953.

5. A Method of Measuring Earth Resistivity, Frank Wenner. *Bulletin*, National Bureau of Standards, Washington, D. C., vol. 12.

6. The Calculation of Resistances to Ground and of Capacitance, H. B. Dwight. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. X, no. 1, 1931.

7. Electric and Magnetic Fields (book), S. S. Attwood. John Wiley & Sons, Inc., New York, N. Y., 1949.

◆

Discussion

G. F. Tagg (Evershed & Vignoles, London, England): The author has set out clearly one of the big difficulties in making accurate ground resistance measurements. In particular he has exposed the fallacy of obtaining more accurate results if the potential point is on the far side of the electrode from the current sink. On the other hand, it is only fair to remember that owing to variations in the resistivity of the soil, the proximity of topological features, and other factors, it is almost impossible to expect a very high degree of accuracy in measurements of this type. This being so, an application of the author's rule for distance may give no more accurate results than a measurement made in the usual way when the current and potential points are placed sufficiently far from the electrode under test for the resistance-distance curve to show a horizontal portion.

A paper of this kind is particularly valuable in drawing attention to the fundamental factors which underly this type of measurement. In dealing with very large distributed grounds it would, unfortunately, be almost impossible to apply the results of a theoretical investigation such as this, and recourse must be had to the practical rule of obtaining a horizontal section on the resistance-distance curve.

E. B. Curdts: Dr. Tagg has very properly emphasized the statement at the beginning of this paper that there is no theoretical substitute for actual field measurements of ground resistance. However, I am sure that we all agree on the desirability of a theoretical basis for all such measurements.

Since such measurements are ordinarily made with alternating or periodically reversing voltages rather than with direct voltages, the mathematical analysis in this paper lacks the rigor that would obtain on the basis of impedance paths rather than purely resistive ones. This may, in some instances, be one of the causes of the difference between theoretical and actual results.

One of the justifications for a theoretical analysis of the problem as given in the paper is the fact that some of the available publications on the subject convey, by implication at least, the idea that horizontal sections of the resistance-distance curves mentioned by Dr. Tagg are to be expected from a theoretical point of view.

The cost of making actual earth resistance measurements on very large ground grids can be appreciable, and anything done to provide a better understanding of the fundamentals involved should tend to reduce such costs.

The Technique and Instrumentation of Low-Impedance Ground Measurement

C. A DUKE
FELLOW AIEE

L. E. SMITH
NONMEMBER AIEE

THE MEASUREMENT of the impedance of grounding systems that have a value of less than one ohm and also residual ground currents flowing at the time of measurement requires special techniques and instrumentation. A summation of the quantities involved is given in this paper, with a description of a practical instrument that has been developed for making this type of measurement.

Measurement and Instrumentation Considerations

The impedance of a large-area grounding system is a vector quantity and can have a relatively large reactive component. Preferably the impedance is measured at system frequency, and the simplest method uses the voltmeter-ammeter arrangement.[1]

The instrument described in this paper is a specialized voltmeter-ammeter that responds to components of system frequency only. The instrument has internal circuits that are used to balance the voltages caused by residual currents. These two features permit ground-impedance measurements under difficult conditions by making negligible the chief sources of erroneous indication. The effects of mutual impedance between the test current and the resulting measured voltage must be avoided. This can be done best by taking the voltage readings along a physical line opposite in direction from the application of test current. If it is impractical to use opposite directions, fair results may be obtained at physical angles of 90 degrees between the current-probe direction and the physical line of voltage determinations. These precautions must be observed for low-impedance ground measurement regardless of the type of instruments used.

The effects of mutual impedance always increase the apparent value of ground impedance. The amount of increase varies with the ground resistivity of the area surrounding the ground system that is measured. As an example, on a representative large ground system when the current and potential test probes were placed in opposite directions, a value of 0.15 ohm was determined. When the probes were placed in the same direction, a value of 0.6 ohm was indicated. As a general rule a minimum of 0.3 ohm will be indicated with probes in the same direction, while the true value of the impedance of the ground system may be as low as 0.03 ohm. These differences may have a large economic value in the design of protection systems for communication facilities that are used at the grounding site. The rise of ground potential must be less than the breakdown of the communication protective gaps, or expensive neutralizing transformers must be installed in each communication line. When the possible ground fault current is very high, the difference between 0.1 ohm and 0.3 ohm for station ground impedance may be the deciding factor in determining whether extensive communication protective equipment is needed.

Determination of True Ground Impedance

The maximum value of voltage obtained divided by the test current in amperes gives the effective ground impedance. This gives the true value, provided the test-current probe location is at a greater distance from the center of the ground than the distance at which the maximum voltage is obtained. If

---

Paper 58-105, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 17, 1957; made available for printing March 19, 1958.

C. A. Duke and L. E. Smith are with the Tennessee Valley Authority, Chattanooga, Tenn.

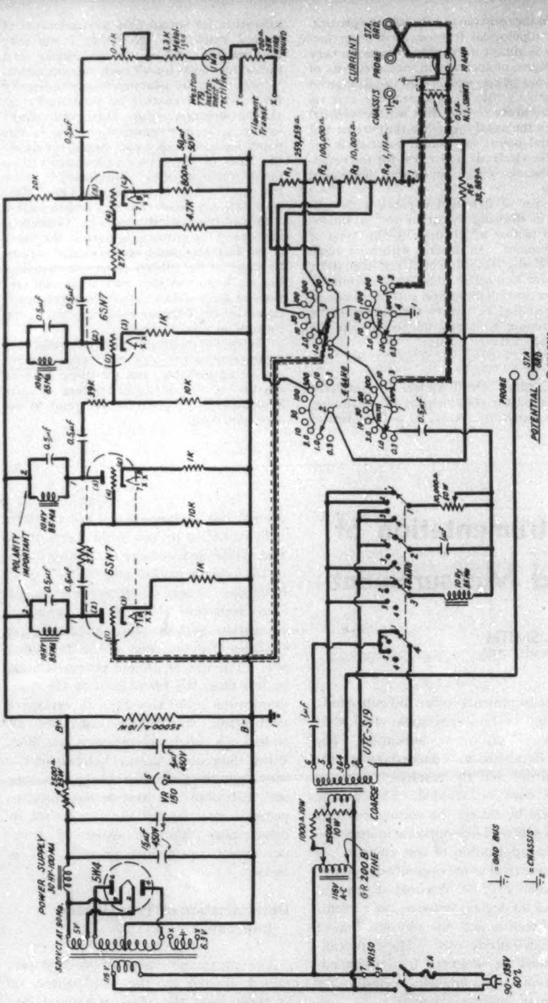


Fig. 1. Schematic wiring diagram of Selvaz

such is not the case, the test-current probe must be moved farther from the grounding system under test to a point of greater distance than that at which the maximum voltage is obtained, and the values must be redetermined. This redetermination is necessary to assure that the point from which the test current enters the ground is effectively at a remote location in relation to the grounding system under test. It is a very rare case that requires a distance of greater than 1,000 feet from the edge of even an extremely large grounding system for the location of the test-current probe, and this distance of 1,000 feet should be chosen for the initial test.

The plot of the ground impedance (volts divided by test amperes) against distance gives the effective potential gradient of the grounding system when the impedance is multiplied by actual or calculated ground-fault current.

Instrumentation Requirements

1. Residual Voltages

A substation grounding system usually has residual currents flowing into the system; these currents produce a corresponding voltage across the impedance of the ground. This voltage usually has fairly prominent harmonics. Also, around transmission lines carrying current, voltages may be induced in the test leads. The effects on the ground-impedance determination of these residual voltages must be made negligible.

2. Sensitivity

Values of less than 0.05 ohm for a large station ground impedance may be encountered, and adequate sensitivity must be provided when using a reasonable test current that has been selected arbitrarily at 10 amperes or less.

Actual Instrument Features

The effects of residual voltages are made small by balancing these residual voltages in the test circuit by introducing an equal but oppositely phased voltage, making the net voltage zero in the test circuit, with zero test current. This can be done only if a single-frequency component is considered, so the voltmeter amplifier is made frequency-selective to prevent interference by harmonics of the power line frequency. The response to the third harmonic is only 1%, with less response to higher harmonics. Even-numbered harmonics are rarely found as residuals, but these are rejected also by the voltmeter amplifier. The response to 120 cycles is 5%. The balancing voltage requirements are met with a 20-volt isolated source that can be varied smoothly from zero volts and can be shifted in phase through all angles by means of a quadrant switch and a continuously variable control for variation within each of the four quadrants.

The sensitivity requirements are met adequately, because the voltmeter has a minimum full-scale rating of 0.3 volt, with provision for scale multipliers so that voltages of at least 120 volts can be read to permit quick checking of accuracy in the field.

Need for Absolute Accuracy Eliminated

Since it is difficult to maintain absolute accuracy in a selective-frequency voltmeter that cannot use an inverse feedback amplifier for stabilization, provision is made for using the same amplifier and instrument to measure test current by measuring the voltage drop across a noninductive shunt. A small ammeter is used to monitor the current while the main instrument is measuring potential; thus the impedance (volts divided by amperes) will be indicated correctly, although the absolute indication of the instrument is in error. By using voltage regulation in the plate-voltage supply for the amplifier, the over-all maximum deviation from absolute readings is made relatively small: in the order of 1% over short periods and in the order of 5% from month to month.

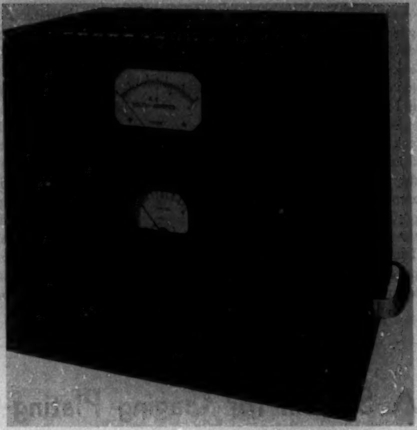


Fig. 2. Instrument used for ground-impedance measurement

Input Circuits

The input impedance to the voltmeter amplifier should be as high as practical so that the temporary potential test probe resistance to ground will have a negligible effect on the indicated voltage. In poor grounding locations the probe resistance may be as high as 3,000 ohms. In the final instrument the input impedance is 300,000 ohms; so no corrections need to be applied for probe resistance.

The terminal arrangements on the final instrument permit separate ground connections for the current and potential circuits. Common or mutual impedances of even short leads must be avoided when very low ground impedances are measured.

Detailed Features of Actual Instrument

The authors have coined the name "Selvaz" for this instrument from the designation term "*sel*ective frequency-*v*olt-*a*mpere-impedance (*Z*) meter."

A schematic wiring diagram is shown as Fig. 1, and the completed instrument is shown as Fig. 2. A simplified schematic of the instrument connections and the test ground connections is shown as Fig. 3. Note that two separate connections are made from the instrument to the station ground.

One of the design features of this instrument is the phase-shifting network shown in the lower left-hand corner of Fig. 1, which, when used with a variable-input-voltage isolation transformer, permits the introduction of a variable-magnitude-variable-phase voltage into the potential-measuring circuit. The residual voltage in the potential circuit can be balanced and made negligible; therefore the effect of the test current may be read directly on the voltmeter scale. The phase-shifting network consists of a quadrant switch that places in the circuit either a capacitor or a reactor with an associated variable resistor, with provision for reversing polarities. The phase shift is determined by the setting of the 10,000-ohm variable resistor. The range of the phase shift with the chosen values is 110 degrees, with an overlap on each of the quadrant ranges. A change of magnitude does not introduce a significant phase shift, nor does a change in the phase setting introduce a significant magnitude change when the circuit constants shown in Fig. 1 are used. It should be noted that the residual voltage is balanced directly before the potential-measuring circuit is involved. This permits changing voltmeter ranges without the necessity of rebalancing the residual voltage. With the circuit constants as given, about 20 volts of balancing voltage is available; this is adequate for any well-designed grounding system.

Proceeding further in the potential-measuring circuit, the 4-gang volt-ampere range switch is used to prevent sneak circuits between the current and potential circuits by means of the first and fourth gang. The second gang on this same switch, by cutting in the shunting 38,889-ohm resistor, gives a multiplier of 3 to the 1-, 10-, and 100-volt ranges; this simplifies the adjustment of the multiplying resistor string.

The selective-frequency amplifier constants were chosen to provide the maximum stability with maximum rejection to harmonics. Some degeneration is provided by not bypassing the cathode resistors in the tuned-plate stages. The reactors in the plate circuit of the first two stages are oriented so that the influence

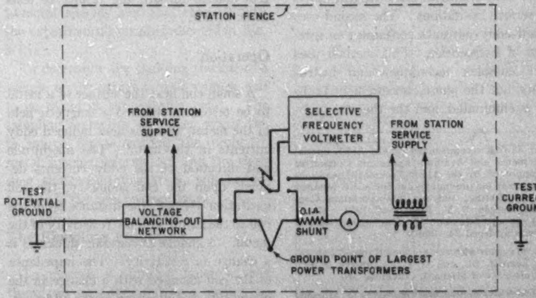

STATION FENCE
FROM STATION SERVICE SUPPLY
SELECTIVE FREQUENCY VOLTMETER
FROM STATION SERVICE SUPPLY
TEST POTENTIAL GROUND
VOLTAGE BALANCING-OUT NETWORK
0.1Ω
SHUNT
A
TEST CURRENT GROUND
GROUND POINT OF LARGEST POWER TRANSFORMERS

Fig. 3. Simplified schematic of instrument connections and test ground connections

of external magnetic fields is made negligible. Other circuits that do not use inductances in the selective-frequency circuits have been tried experimentally but have been found inferior to the one shown in Fig. 1.

With all the possibilities of extraneous voltages appearing in the amplifier circuit from either stray capacitances or stray flux fields, it is impractical to get a true zero response on the instrument when the potential input is short-circuited or when the range switch is set on "amperes." The simplest solution to this problem is provided in the final indicating instrument circuit, where a line frequency voltage is introduced in series with the indicating instrument. This voltage appears across the 100-ohm potentiometer-type rheostat provided from the center-tapped 6.3-volt winding of the amplifier power transformer. In operation the electrical zero of the indicating instrument is brought to true zero by means of this rheostat, while the range switch is set on amperes. This adjustment is not brought out on the panel, since usually a single setting is used on each instrument. Any internal residual voltage that may be present in the "volts" measuring position will be balanced manually and will not affect the measurement.

Conclusions

Five of these instruments have been in service for several years on the system of the Tennessee Valley Authority and have given very satisfactory results under difficult measurement conditions. The design features were developed because of specific requirements involved in making impedance measurements on extensive grounding systems.

Reference

1. THE MEASUREMENT OF THE IMPEDANCE OF LARGE SUBSTATION GROUNDS, L. E. Smith. *AIEE CP 51-147* (available on request).

A System for Gaging Plating Thickness

R. G. MYERS
ASSOCIATE MEMBER AIEE

D. L. WAIDELICH
FELLOW AIEE

THE NONDESTRUCTIVE measurement of cladding thickness when the cladding metal and the base metal are nonmagnetic has presented a major problem for a number of years. There have been several different approaches to the problem such as ultrasonics, back scattering of beta or gamma rays, and eddy currents. At present, the eddy-current method appears to have the greatest chance for success. Several eddy-current instruments have been developed, but there is still a need for improved instruments.

This eddy-current method utilized two different techniques in application. The first[1–2] employs sinusoidal eddy currents of a single frequency. This method uses comparatively simple, well-known techniques and instruments, but inaccurate phase measurements and thermal drifts are serious limitations. The second uses pulsed eddy currents containing a spectrum of frequencies. This method uses more complex techniques and instruments, but the phase measurement problem is eliminated, and the thermal problem created by current in the probe coil is minimized.

In the first pulsed eddy-current instrument,[3–5] the cladding thickness was read from the screen of a cathode-ray oscillograph. This instrument suffered from these disadvantages: it was not useful for continuously scanning a metal, and the coating thickness information was presented on the face of an oscilloscope and could not be recorded by a recording voltmeter. A second instrument[6] overcame these difficulties to some extent by the use of photocells placed on the screen of the oscilloscope. It was desirable, however, to eliminate the use of the oscilloscope and photocells if possible, and a new system for transferring the information to a recorder was developed. It is the purpose of this paper to describe this system and to present the results of tests made with it.

Operation

A small coil near the surface of a metal to be tested established a magnetic field in the metal, and this field induced eddy currents in the metal. The magnitude and direction of the eddy currents depends upon the coil geometry, the coil excitation voltage, the distance from the coil to the metal, and the resistivity of the metal. A change in cladding thickness is a change in resistivity. The impedance of the coil changed with a change in the magnitude and direction of the eddy currents.

A small ferrite-cored coil with its axis perpendicular to the surface of the metal was used to project the field into the metal. The coil windings were concentrated near the end of the probe to obtain maximum sensitivity to a given change in coating thickness. A cross section of the probe is shown in Fig. 1. When the cladding thickness changed, the impedance of the coil changed. To obtain the cladding thickness information, this change in coil impedance was measured with the aid of a bridge circuit.

The bridge circuit in Fig. 2 was excited by discharging a charged capacitor through the thyratron. When the input pulse from the pulse generator triggered the thyratron the 1,000-$\mu\mu f$ (micromicrofarad) capacitor discharged through a pulse-shaping network and the primary of the transformer. This provided a bridge excitation pulse shaped like the positive half of a sinusoidal wave. The secondary of the transformer was loosely coupled to the primary and shielded from the primary.

The pulse projected into the coated metal should meet certain length requirements. The length is directly related to the basic time.[3]

$$T=t^2\mu_1\sigma_1$$

where t is the coating thickness. Representative values of T are given in Table I. For best results, the input pulse from the probe should be approximately five times

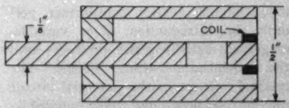


Fig. 1. Probe cross section

Paper 58-933, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted January 17, 1958; made available for printing April 24, 1958.

R. G. MYERS is with the Argonne National Laboratory, Lemont, Ill., and D. L. WAIDELICH is with the University of Missouri, Columbia, Mo.

This work was done at the University of Missouri Engineering Experiment Station, under a contract sponsored by the Argonne National Laboratory.

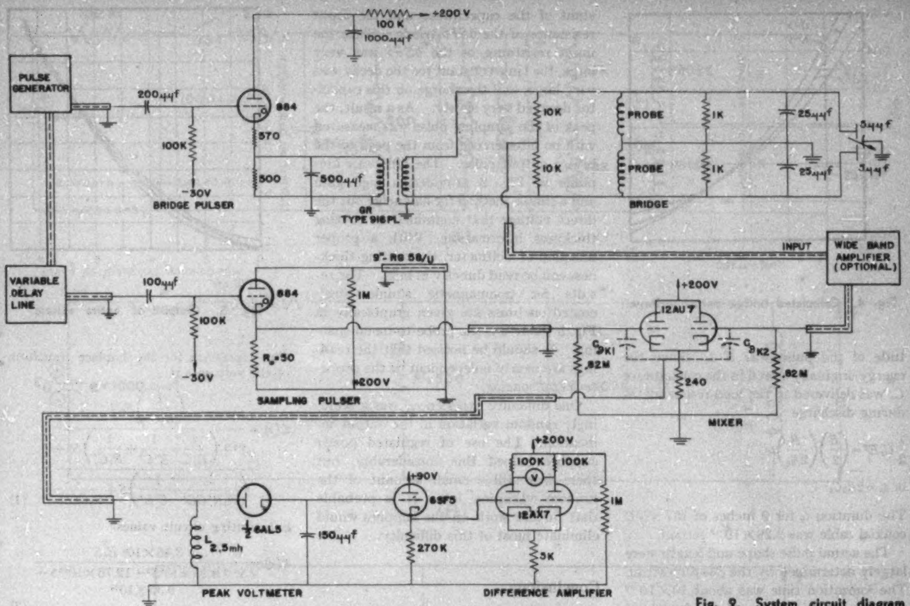

PULSE GENERATOR
200μμf
100K
-30V
BRIDGE PULSER
884
570
500
500μμf
GR TYPE 916PL
100 K
1000μμf
+200 V
10K
10K
PROBE
PROBE
BRIDGE
1K
1K
25μμf
25μμf
5μμf
5μμf
INPUT
WIDE BAND AMPLIFIER (OPTIONAL)
VARIABLE DELAY LINE
100μμf
100K
-30V
884
1M
9"- RG 58/U
R₀=50
+200V
SAMPLING PULSER
+200V
12AU7
C_{gK1}
C_{gK2}
.82M
.82M
240
MIXER
L 2.5mh
½ 6AL5
150μμf
PEAK VOLTMETER
+90V
6SF5
270K
+200V
100K
100K
V
12AX7
1M
5K
DIFFERENCE AMPLIFIER

Fig. 2. System circuit diagram

T. From Table I this would necessitate an extreme range of pulse lengths. Some compromise is necessary unless only one coating thickness and coating metal is to be used.

If a slight unbalance is present, an unbalanced voltage exists at the bridge output terminals. A standard sample of metal was placed on one probe, and another sample of metal was placed on the other. Balancing adjustments on the bridge were made, and the bridge output voltage was nearly zero. An unbalance was added by changing the test probe air gap, and the output voltage pulse was as indicated in Fig. 3(A). One major difficulty was variation of the bridge output voltage with probe-to-metal spacing. It was discovered experimentally that one point on the bridge output pulse did not vary in amplitude with the probe-to-metal spacing as indicated in Fig. 3(B). This point will be called the crossing point. The amplitude of the crossing point does vary with the cladding thickness.[6]

Table I. Values of T

Coating Metal	T Microseconds Coating Thickness 5 Mils	25 Mils
Aluminum	0.75	18.8
Zirconium	0.045	1.12
347 stainless steel	28.1	702.0

An analysis of the bridge circuit was made to show that the crossing point does exist. The details of the analysis are given in the Appendix, and the results are shown in Fig. 4, where e_0 is the bridge output voltage and K' is a constant which does not change the shape of the curves. There was one point on the calculated curves that did not vary with the probe-to-metal spacing and this checked with the experimental results indicated in Fig. 3(B).

To determine the cladding thickness, a sample of the bridge output voltage had to be taken at the crossing point. The sampling technique, in other words the technique for separating the probe-to-metal spacing information and the cladding thickness information, will be described in the next two sections.

The Sampling Pulse

To generate a sampling pulse, an *884* thyratron was operated as a switch to discharge a charged transmission line through the characteristic impedance of line as indicated in Fig. 2. The pulse from the bridge circuit and the sampling pulse were synchronized by firing both thyratrons with one pulse generator. The sampling pulse was positioned in time by a variable delay line in the grid circuit of the thyratron.

The duration t_0 of the sampling pulse was calculated from the value of the total storage capacitance C_s and the characteristic impedance R_0 of the line. The network was assumed lossless and was charged to a voltage E. The ampli-

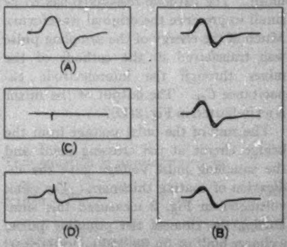


Fig. 3. Waveforms from cathode-ray oscilloscope

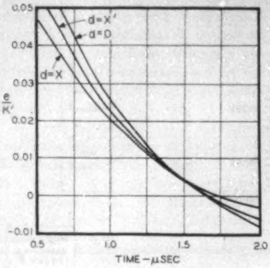


Fig. 4. Calculated bridge output voltage

tude of the pulse was $E/2$. Since the energy originally stored in the capacitance C_s was delivered to the load resistance R_0 during discharge

$$\frac{1}{2} C_sE^2=\left(\frac{E}{2}\right)\left(\frac{E}{2R_0}\right)t_0$$

or $t_0=2R_0C_s$

The duration t_0 for 9 inches of *RG-58/U* coaxial cable was 2.29×10^{-9} second.

The actual pulse shape and length were largely determined by the *884* thyratron. The ionization time was about 10×10^{-9} second and this increased the length of the pulse. The deionization time of the *884* had a very small effect on the pulse length. The transmission line discharged very rapidly, and when the plate voltage dropped to zero, the tube extinguished. Fig. 3(C) shows the sampling pulse. The pulse length was about 15×10^{-9} second and the amplitude was approximately 15 volts.

Mixer and Voltmeter

The mixer in Fig. 2 summed the bridge output pulse and the sampling pulse. The high input impedance reduced the interaction of the two pulses to a minimum. The cathode resistance has to be small to preserve the original waveforms. Much of the energy of the sampling pulse was transferred to the cathode of the mixer through the interelectrode capacitance C_{gk}. The output of the mixer is as indicated in Fig. 3(D).

The sum of the pulse voltage from the bridge circuit at the crossing point and the sampling pulse voltage gave the indication of coating thickness. The peak voltmeter in Fig. 2 measured this sum. During the time of the sampling pulse, a charge built up on the 150-μμf capacitor. During the off part of the cycle the charge decayed at a rate given by the time constant of the capacitance and the input resistance of the *6SF5* triode. Since the input resistance of the *6SF5* was very large, the time constant for the decay was very large, and the charge on this capacitor decayed very slowly. As a result, the peak of the sampling pulse was measured with no interference from the peak of the bridge output pulse. The difference amplifier in Fig. 2 provided a high gain and a simple method for balancing out the direct voltage that contained no coating thickness information. With a proper selection of voltmeter, the coating thickness can be read directly in mils. The results for nonmagnetic stainless-steel-coated on brass are given graphically in Fig. 5 where d is the probe-to-metal spacing. It should be noticed that the readings are nearly independent of the probe-to-metal spacing.

One difficulty experienced was a seemingly random variation in the output indication. The use of regulated power supplies helped this considerably, but there was still a small amount of the random variation left. It is probable that further work on the supplies would eliminate most of this difficulty.

Conclusions

The calculated bridge output voltage indicated the presence of the crossing point. The major error in the calculation was introduced by the probe characteristic assumptions. A true equivalent circuit for the probe has not been found, and this will have to be done before exact calculations can be made.

This eddy-current instrument measures cladding thickness fairly independent of probe-to-metal spacing. Pulsed eddy currents have one major advantage over sinusoidal eddy currents: a much higher peak current can be applied to the probe coil without heating the coil if the proper pulse length and repetition rate have been selected. The one major disadvantage of this pulse system is that the pulses must be very stable in amplitude and time.

Appendix

Bridge Circuit Analysis

The equivalent circuit of the bridge circuit and bridge circuit pulser is shown in Fig. 6(A). The voltage e was calculated by replacing the transformer and the bridge with an equivalent circuit which was found by measuring the input impedance at the transformer primary. Then the equivalent circuit is as shown in Fig. 6(B).

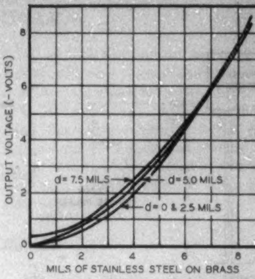


Fig. 5. Output of entire system

The expression for the Laplace transform of the voltage e is

$$\mathfrak{L}(e)=\frac{\left(\frac{1}{R_1C'}\right)E_0S}{S^3+\left(\frac{1}{R_1C}+\frac{1}{R'C'}+\frac{1}{R_1C'}\right)S^2+\left(\frac{1}{R_1R'CC'}+\frac{1}{C'L'}\right)S+\frac{1}{R_1CC'L'}} \quad (1)$$

Substituting circuit values

$$\mathfrak{L}(e)=\frac{3.46\times10^6\,E_0S}{S^3+9.15\times10^6S^2+12.76\times10^{12}S+9.35\times10^{18}} \quad (2)$$

The Laplace transform of the primary current is

$$\mathfrak{L}(i_p)=\frac{1}{L'S}\,\mathfrak{L}(e) \quad (3)$$

After substituting equation 2 into equation 3 and taking the inverse Laplace transform, the primary current i_p is

$$i_p=1.67\times10^{-3}E_0\epsilon^{-0.755\times10^6t}\times \sin(0.806\times10^6t-6.8°) \quad (4)$$

for $t\geqq0.3$ μsec (microsecond). The Laplace transform for the approximate primary current i_p is:

$$\mathfrak{L}(i_p)=\frac{-0.197\times10^{-3}E_0S+1.19\times10^3E_0}{[(S+0.755\times10^6)^2+(0.806\times10^6)]^2} \quad (5)$$

The resistances R_1 and R_2 were assumed small enough to be neglected. The 10-kilohm resistances may be neglected, because they are large compared to the 1-kilohm resistances. The large input resistance R_d of the mixer was neglected.

With the bridge balanced, the voltage equation for the secondary is

$$0=M\frac{di_p}{dt}+L_s\frac{di_2}{dt}+i_2Z \quad (6)$$

where Z is the entire bridge impedance, and

$$Z(s)=\frac{2LRS}{RCLS^2+LS+R} \quad (7)$$

The Laplace transform of equation 6 was solved for the Laplace transform of the

Vol.
Pt. 1
1958

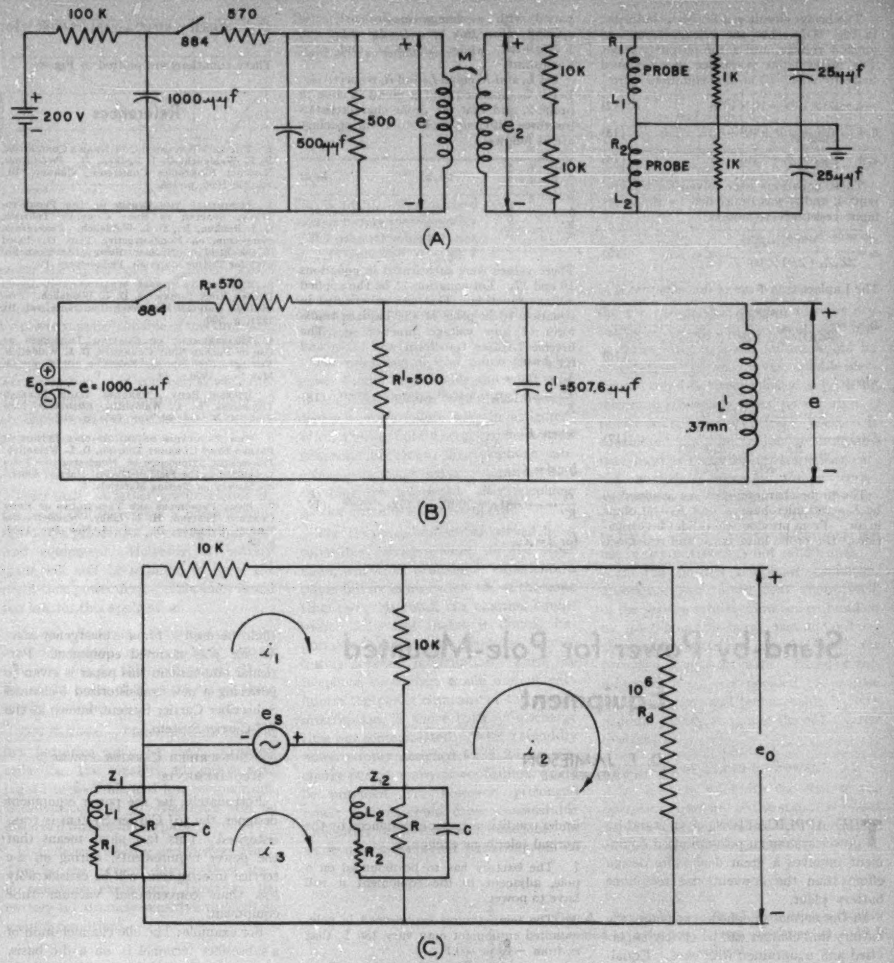

100 K
570
884
+
200 V
−
1000μμf
500μμf
500
e
M
e₂
10K
10K
R₁
L₁
PROBE
R₂
L₂
PROBE
1K
1K
25μμf
25μμf
(A)
R₁=570
884
E₀
c=1000μμf
R′=500
C′=507.6μμf
L′
.37mh
e
(B)
10 K
i₁
10K
e_s
Z₁
L₁
R
R₁
C
i₃
Z₂
L₂
R
R₂
C
i₂
10⁶
R_d
e₀
(C)

Fig. 6. Equivalent circuit

A—Of bridge pulser and bridge
B—For calculating voltage across transformer primary
C—Of the bridge

secondary current. Multiply by $Z(s)$ to obtain

$$\mathcal{L}(e_2)=\frac{-MS\mathcal{L}(i_p)Z(s)}{L_sS+Z(s)} \tag{8}$$

The circuit values were substituted in equation 5 and the inverse Laplace transform was used to obtain

$$\frac{e_2}{K10^{-9}}=9.8\epsilon^{-0.756\times10^{6}t}\sin(0.806\times10^{6}t+46.6^{\circ})+88^{-6.45\times10^{6}t}-1630^{-31.6\times10^{6}t} \tag{9}$$

where

$$K=0.394\times10^{-3}\frac{ME_0}{L_sC}$$

This equation is approximated by

$$\frac{e_2}{K10^{-9}}=45\epsilon^{-0.9\times10^{6}t} \tag{10}$$

or 0.5 μsec ≤ t ≤ 2.0 μsec. The Laplace transform of equation 10 is

$$\frac{\mathcal{L}(e_2)}{K10^{-9}}=\frac{45}{S+0.9\times10^{6}} \tag{11}$$

The bridge circuit was drawn as indicated in Fig. 6(C). Let the voltage e_s be any applied voltage, and e_0 the output voltage. The mixer input resistance was assumed to be 10^6 ohms. The loop equations were

$$-e_s=20\times10^3 i_1-10\times10^3 i_2 \qquad (12)$$

$$0=-10^4 i_1+(10^4+10^6+Z_2)i_2-Z_2 i_3 \qquad (13)$$

$$e_s=-Z_2 i_2+(Z_1+Z_2)i_3 \qquad (14)$$

These equations were solved for the current i_3, and i_3 was multiplied by the mixer input resistance to obtain

$$e_0=\frac{10^6(Z_2-Z_1)e_s}{2Z_1Z_2+2.01\times10^6(Z_1+Z_2)} \qquad (15)$$

The Laplace transform of this equation is

$$\mathfrak{L}(e_0)=\frac{10^6[Z_2(s)-Z_1(s)]\mathfrak{L}(e_s)}{2Z_1(s)Z_2(s)+2.01\times10^6[Z_1(s)+Z_2(s)]} \qquad (16)$$

where

$$Z_i(s)=\frac{\frac{1}{C}S+\frac{R_i}{L_i}}{S^2+\frac{R_i}{L_i}+\frac{1}{RC}S+\frac{R_i+R}{CL_i}} \qquad (17)$$

The probe characteristics are assumed to be $L_0=200$ microhenrys, and $R_0=50$ ohms in air. From previous sinusoidal investigations,[1] the probe inductance and resistance varied with a change in probe-to-metal spacing according to $L=L_0-\Delta L$, and $R=R_0+\Delta R$, where $\Delta R=K\Delta L$ and K is a constant.

Let L_1 and R_1 equal L_0 and R_0 respectively. Let it be assumed that a metal is close to probe 2, and that the probe characteristics for three different probe-to-metal spacings are as follows:

d	$L_2(\mu h)$	$R_2(\Omega)$
0	120	80
X'	130	75
X	140	70

These values were substituted in equations 16 and 17. Let equation 11 be the applied voltage function. This was substituted in equation 16 in place of the Laplace transform of any voltage function e_s. The inverse Laplace transform was taken and for $d=0$,

$$\frac{e_0}{K'}=-0.105\epsilon^{-0.403\times10^6t}+0.236\epsilon^{-0.9\times10^6t} \qquad (18)$$

where $K'=\frac{-45}{K10^{-6}}$

for $d=X'$

$$\frac{e_0}{K'}=-0.081\epsilon^{-0.373\times10^6t}+0.192\epsilon^{-19\times10^6t} \qquad (19)$$

for $d=X$

$$\frac{e_0}{K'}=-0.059\epsilon^{-0.352\times10^6t}+0.151\epsilon^{-0.9\times10^6t} \qquad (20)$$

These equations are plotted in Fig. 4.

References

1. The Impedance of a Coil Near a Conductor, D. L. Waidelich, C. J. Renken, Jr. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956, p. 188.

2. Minimizing the Effect of the Probe-to-Metal Spacing in Eddy Current Testing, C. J. Renken, Jr., D. L. Waidelich. *Proceedings*, Symposium on Nondestructive Tests Developed in the Field of Nuclear Energy, American Society for Testing Materials, Philadelphia, Pa.

3. Coating Thickness Measurements Using Pulsed Eddy Currents, D. L. Waidelich. *Proceedings*, National Electronics Conference, vol. 10, 1954, p. 500.

4. Measurement of Coating Thickness by Use of Pulsed Eddy Currents, D. L. Waidelich. *Nondestructive Testing*, Evanston, Ill., vol. 14, May-June, 1956,p. 14.

5. Pulsed Eddy Currents Gage Plating Thickness, D. L. Waidelich. *Electronics*, New York, N. Y., vol. 28, Nov. 1955, p. 146.

6. The Reduction of the Spacing Effect in Pulsed Eddy Current Testing, D. L. Waidelich. *Proceedings*, Symposium on Nondestructive Tests Developed in the Field of Nuclear Energy, American Society for Testing Materials.

7. Basic Principles and Techniques of Eddy Current Testing, H. L. Libby. *Nondestructive Testing*, Evanston, Ill., vol. 14, Nov.-Dec. 1956, p. 12.

Stand-by Power for Pole-Mounted Equipment

D. F. JAMIESON
NONMEMBER AIEE

THE APPLICATION of a stand-by power system to pole-mounted equipment involves a great deal more design effort than the conventional telephone battery plant.

In the normal telephone exchange, the battery and charger can be closely supervised and maintained with ease. Equalizing charges can be accomplished at frequent intervals, and specific gravity readings can be taken, over a period of time, without difficulty. The temperature variations experienced by a normal telephone exchange battery plant are not more than ±30 F (degrees Fahrenheit) from room temperature and published discharge curves predict the battery behavior during an a-c power interruption. All of these conditions help to keep the battery at its top efficiency.

To power pole-mounted equipment, the stand-by battery plant has to perform under conditions never experienced by the normal telephone exchange:

1. The battery has to be mounted on a pole, adjacent to the equipment it will have to power.
2. The temperatures experienced in pole-mounted equipment may vary 180 F, that is, from −40 to 140 F.
3. Servicing of the battery plant will have to be kept to a minimum, due to the inaccessibility of the equipment located at widely scattered points, many miles from the central office.
4. Water additions to the battery will not be easily accomplished in most cases and should not be required except on a semiannual basis.
5. Available charge and discharge curves do not give any indication of the battery's characteristics when operating in this environment.

These problems started an investigation into the type of stand-by power which could be used to provide emergency service for pole-mounted equipment. Particular emphasis in this paper is given to powering a new transistorized 5-channel Subscriber Carrier System, known as the *561* Carrier System.

561 Subscriber Carrier Power Requirements

Fortunately, for the power equipment designer, the *561* Carrier System is transistorized. This fact alone means that the power requirements, during an a-c service interruption, will be considerably less than conventional vacuum tube equipment.

For example: The idle channel drain of a subscriber terminal is, on a d-c basis, only 5 watts. When the channel is in use, its power drain increases to 10 watts.

Paper 58-851, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 18, 1958; made available for printing April 11, 1958.

D. F. Jamieson is with Stromberg-Carlson Company, a Division of General Dynamics Corporation, Rochester, N. Y.

The author wishes to acknowledge the assistance of J. E. Ross, who supervised the mechanical design; Henry Jenson of the C & D Battery Company for the use of data on the lead calcium batteries; and R. L. Layburn, S. B. Weiner and H. C. Behnk of Stromberg-Carlson who assisted in the preparation of this paper.

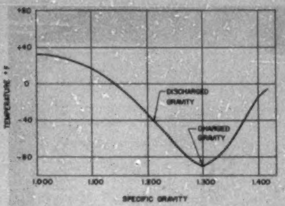


Fig. 1 (left). Freezing points of aqueous solutions of sulphuric acid at various gravities

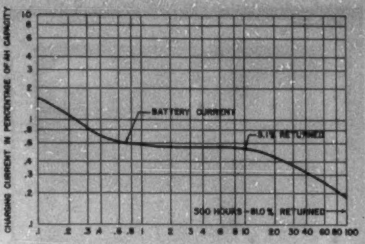


Fig. 3 (right). Typical lead calcium battery specific gravity 1.300 constant voltage charge at 2.35 volts/cell at −40 F

Most of this additional energy goes to power the subscriber's telephone. Peak power requirements during the signaling interval does not exceed 25 watts. Stand-by power supplies capable of meeting such low power requirements are numerous. Many types of stand-by power supplies were investigated in connection with this development. The most promising types are listed, with a short description of their advantages or limitations.

TYPE OF STAND-BY POWER SUPPLIES

Solar Cells. A rather new development, the Solar Cell represents a possible solution to the problem of powering transistorized equipment. However, a battery plant will still be required to cover the night-time power drain. Efficiency is still too low for this application.

Atomic Batteries. Several types are available today; however, their output is too low even for this application. Shelf-life of the order of 20 years makes them highly desirable.

Dry Batteries. Many different types of dry batteries could be used. Mercury cells offer the greatest advantage with regard to shelf-life and low temperature. Disadvantages of dry batteries will be discussed further in this paper.

A-C Battery Charger and Battery Plant. This is the present-day method used in telephone exchanges. However, the battery is extremely sensitive to environmental conditions.

SERVICE REQUIREMENTS FOR STAND-BY POWER

The questions of how much stand-by power should be available and what type of telephone service would be required during a power failure were investigated. When a power failure occurs in the normal telephone exchange, the telephone subscriber is in many instances unaware that anything has happened. His telephone service continues uninterrupted.

The telephone subscriber served by a subscriber carrier system, on the other hand, will in all probability experience a power failure in his own home at the same time power is cutoff the channel equipment. This fact makes it doubly important to maintain telephone service during a power interruption. Since the telephone subscribers would undoubtedly inform the power company of the service interruption, a short battery discharge time was contemplated. With a stand-by power supply designed for a 2-hour discharge period, many power failures would be provided for; however, prolonged power failures could cause considerable difficulties with customer relations. A short discharge period gave impetus to the use of one-shot devices such as dry batteries. The replacement of the batteries after a power failure would be necessary in order to provide reliable service and could be accomplished only if the telephone company was aware that a power failure had occurred. When it is considered that the telephone company may have as many as 100 subscriber carrier channels in operation, the service requirements are tremendous and the battery replacement cost is extremely expensive. Under these considerations, the use of dry batteries is not practicable.

In the normal telephone exchange, experience has shown that about 98% of the service interruption can be handled by an 8-hour discharge period, and rechargeable batteries can be used for periods as long as 20 years. Based on this information, it was decided to provide an a-c charger and rechargeable battery of high quality to power the *561* Carrier System.

ECONOMICS OF STAND-BY POWER

In order to minimize the cost to the operating company of the stand-by power supply, the requirements were established on the basis of two groups:

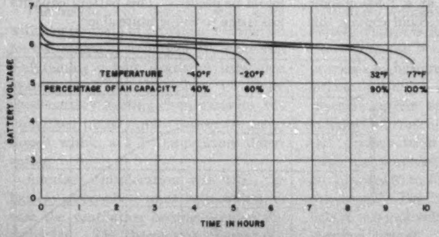


Fig. 2. Typical lead calcium battery specific gravity 1.300 constant current discharge to 1.84 volt/cell

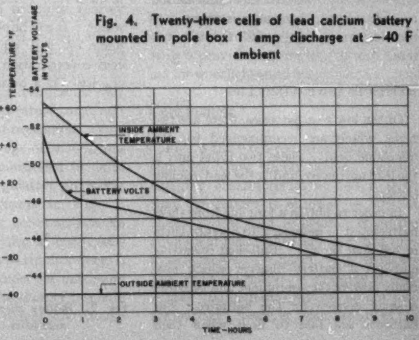


Fig. 4. Twenty-three cells of lead calcium battery mounted in pole box 1 amp discharge at −40 F ambient

Fig. 5 (left). Laboratory setup used in environmental testing of batteries, charger, and insulated housing

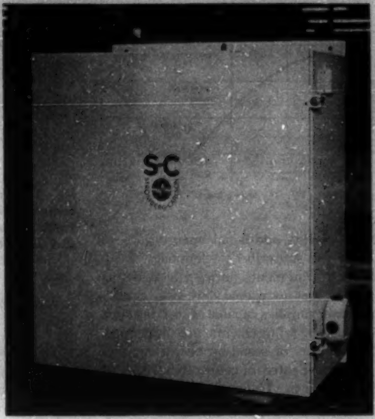


Fig. 6 (right). Engineering model of stand-by power supply housing

Group 1. A battery and charger capable of powering five subscriber terminals for an 8-hour discharge period.

Group 2. A battery and charger capable of powering one subscriber terminal (with provision for growth to three terminals) for an 8-hour discharge period.

This grouping enabled the designer to take advantage of types of batteries already in high production and make a considerable saving in cost. The grouping of the power requirements also indicated that for a single terminal power source, the cost analysis should consider the growth factor. This is necessary because the single terminal itself could not be expected to support the standby power supply. In special cases where standby power is required for only the single channel, it would be necessary to plan the installation with this growth in mind. This can be accomplished easily with the *561* Carrier by concentration of the terminals. For further information on this subject, the reader is referred to the literature on the *561* Carrier System.[1-3]

BATTERIES AND BATTERY LIFE

Of the low cost batteries, there are two types which are recommended for the telephone field. These two types are: 1. the lead antimony, and 2. the lead calcium battery.

In order to obtain a long life from these batteries, it is necessary to operate them in full float service. Full-float operation employs a battery charger which normally powers the equipment and provides a small current for the batteries. The batteries are said to float when their internal losses are compensated for and the battery voltage is raised to the gassing voltage. In this operation, the charging current is very low and, as a result, long life can be expected from the batteries.

Once the basic power system was decided upon, the problems of low temperature operation of the lead antimony battery and the lead calcium battery were considered.

LOW TEMPERATURE OPERATION OF LEAD ACID BATTERIES

The main problems of operating batteries at low temperature are:

1. The battery once it is discharged will freeze.
2. The battery ampere-hour capacity is reduced.
3. Recharging the battery after an 8-hour discharge period requires an extremely long period of time.

Fig. 1 shows the freezing points of various solutions of sulphuric acid and water. It will be observed that the lowest point on this curve occurs when the acid concentration is at a 1.300 specific gravity. If a battery could contain this high-gravity acid, it would freeze at approximately −85 F. A lead antimony battery containing this high acid concentration would have a very short life. The lead calcium battery, on the other hand, could handle this concentration without any ill effects. This would mean that a lead calcium battery could not freeze, buckle its plates, crack its container, or the like, at a temperature of −40 F since its discharged gravity would still be higher than 1.240 after an 8-hour discharge period. Going to a higher acid concentration than 1.300 would allow the discharged gravity to fall to the 1.300 point. Higher acid concentration would, however, shorten the life of even a lead calcium battery. Thus, a compromise was made and the battery was established to be composed of lead calcium cells at a 1.300 specific gravity.

Discharge curves of a 1.300 gravity lead calcium battery shown in Fig. 2 indicates that in order to provide the power requirements over the low temperature range, a battery of 60% more capacity would be needed. This battery capacity loss is due to temperature alone.

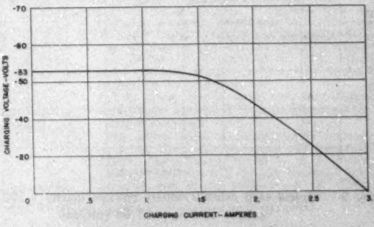


Fig. 7. Battery charger characteristics

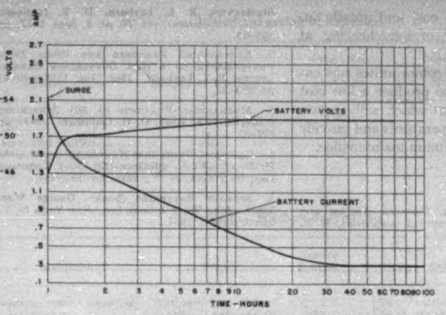


Fig. 8. Twenty-three cells of lead calcium battery mounted in pole box charge characteristic at 140 F

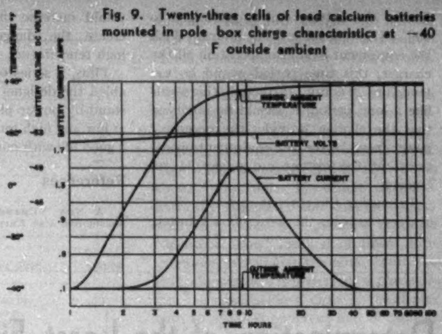


Fig. 9. Twenty-three cells of lead calcium batteries mounted in pole box charge characteristics at −40 F outside ambient

The charging characteristics of the lead calcium cell at a temperature of −40 F, Fig. 3, indicates that approximately 500 hours will be required to return the battery to 81% of its capacity. Since a stand-by power source requiring 500 hours to recharge would be useless for the *561* application, a source of heat would have to be provided.

Battery Housing Design

Because it is necessary to provide heat for the batteries in low temperature operation, the battery housing must contain a heater. In order to reduce the size and power drain of the heater, it is necessary to equip the housing with a thermostat and insulate the box to reduce the heat loss. Thus, when a power failure occurs, the heat in the battery housing plus the heat stored in the batteries will provide a good 8-hour discharge characteristic; see Fig. 4.

The battery problem might seem to be solved merely by sealing and heating the battery housing. This proved misleading. Those familiar with lead acid batteries know that hydrogen gas is given off during the charging and floating periods and ventilation is required to prevent an explosion.

The mechanical designer was faced with the problem of providing ventilation for the escape of hydrogen gas while still maintaining a high degree of insulation against heat loss. In order to insure good battery floating characteristics, the individual battery cells had to be maintained within a 5 F temperature differential.

Another consideration was that the battery atmosphere must be sparkless, in case the ventilation became accidentally restricted. The thermostat, charger disconnect relay, and all possible sparking points are, therefore, enclosed. Ventilation for the housing is accomplished by small holes located in the bottom, and at the top side of the box; see Figs. 5 and 6. Because of the vent hole placement, gas escape is assured and the heat loss is limited to that through the top vent.

The box itself is insulated with fiber glass insulation with an aluminum reflecting surface. The heater, a strip type, insulated with Teflon to prevent acid damage, is located in the bottom of the box. The batteries are arranged above the heater on a perforated shelf in two steps, and on a solid shelf near the top with the charger.

This arrangement is such that air through the bottom vent is heated and is forced to circulate around the batteries. The air is forced, by its own rising characteristics when it hits the solid shelf, to circulate throughout the box. With this design, it is possible to maintain the battery temperature differential to within 5 F.

Power System Layout

The power plant consists of 23 cells of lead calcium battery, designed to float at 2.30±0.05 volts per cell. The battery charger will parallel the battery and hold the battery voltage at 52.9±1 volt. The battery charger is designed to operate over the −40 to 140 F temperature range and provide a high degree of regulation. Variations on a-c line voltage of ±10% and charging current are controlled by both voltage and current regulators producing the characteristics shown in Fig. 7. Taps are provided to accommodate an a-c range of 95 to 130 volts, 60 cps (cycles per second). The 120-cps ripple voltage is held to 300 millivolts under the maximum charging current. The charger response time to changing loads is held slow enough so as not to produce noise in the *561* Carrier System.

Power Plant Maintenance

The problems of maintenance for both the battery and charger were paramount in the mind of the designer. The battery is easily accessible both for specific gravity readings and water additions. The charger circuitry employs the use of long-life components. Silicon power diodes are used to remove the problem of aging adjustments on rectifier stacks and to reduce the size of the units. Either magnetic or transistor regulation was recommended in order to eliminate the maintenance difficulties of vacuum tube control.

The charger, at present, has one adjustment for float voltage; however, this adjustment will be removed in later design and will eliminate the possibility of overcharging due to improper voltage measurement or adjustment. The use of zener diodes for reference elements has greatly reduced the variation of voltage with temperature and will allow the charger circuitry to be set at the factory. Water addition to the batteries is usually not necessary, except in extremely hot climatic conditions. An a-c accessory strip is provided within the battery housing for worklights and instruments.

Conclusions

Figs. 8 and 9 represent the results of the stand-by power supply design. Battery charging characteristics at 140 F show that the battery is returned to 80% of charge in less than 10 hours. When the equipment load is also included, this time is extended to 20 hours. Fig. 9 shows that when the battery is charging at −40 F only 100 milliamperes can be put into the battery; however, with the aid of the heater, the battery can be recharged to 80% of its capacity in 15 hours. With the equipment load also operating off the charger, this time period would be extended to 20 hours. Although, this seems like a long period, it should be observed that the charger provides the necessary power to make the *561* equipment operational at the instant a-c power is reapplied.

Increasing the current capacity of the charger, to bring the battery back faster, would increase the cost and greatly increase the danger of overcharging at high temperature.

Thus, a series of compromises has enabled the designer to produce a low cost stand-by power plant which will operate at low and high temperatures and provide a good life with minimum maintenance.

References

1. A NEW 5-CHANNEL CARRIER SYSTEM FOR SUBSCRIBER USE EMPLOYING TRANSISTORS, B. R. Stachiewicz, R. L. Layburn, D. F. Jamieson. *AIEE Transactions*, vol. 76, pt. I, Sept. 1957, pp. 385–90.

2. ENGINEERING FEATURES AND FIELD TRIAL PERFORMANCE OF A NEW SUBSCRIBER CARRIER SYSTEM, R. L. Layburn. *Ibid.* (Jan. 1958 section), pp. 681–87.

3. MECHANICAL FEATURES OF 561 SUBSCRIBER CARRIER, J. E. Ross, O. H. Goldman. *Ibid.*, pp. 677–81.

4. CHEMICAL ENGINEERS HANDBOOK (book), John H. Perry, editor. McGraw-Hill Book Company, Inc., New York, N. Y., 1951, pp. 184, 281.

5. STORAGE BATTERIES (book), George Wood Vinal. John Wiley & Sons, Inc., New York, N. Y., fourth edition, 1951.

Determination of the Least First Cost of a Cable-Carrier System

ROMAN KRZYCZKOWSKI
NONMEMBER AIEE

IN THE MIDDLE of 1956 a study was made of the economics of providing a cable-carrier system along one of the principal railway routes in England. The route was to be electrified by a 25-kv 50-cps single-phase system which made impossible the continued use of existing overhead telephone lines.

A clean field for complete modernization of the telecommunication system was thus opened. As a result, the technical necessity of introducing suitably screened cable was coupled with the opportunity of providing a substantial increase in general communication facilities, so essential for the best use of the improved efficiency of an electrified railway system.

This paper describes the method adopted in the part of the study devoted to the determination of the Least First Cost of the proposed system. Some of the results and conclusions arrived at in the study are quoted.

Problem and Assumptions

Cable-carrier communication was required along a 160-mile long route. The initial capacity of this system was to be approximately 100 telephone channels and the ultimate capacity in the neighborhood of 400.

The broad examination of the problem in the light of past experience showed that it could be limited to a comparison of implementing required connections by either of the following methods:

1. 12-channel, 2-wire
2. 24-channel, 2-wire
3. Single tube coaxial

In examination of postulated carrier-cable systems the following elements have had to be taken into account:

1. Cable and its method of installation
2. Terminal equipment
3. Line equipment
4. Power feeds

Appropriate combination of these elements gave an optimum economic solution.

It was assumed that irrespective of the trunk circuits to be provided by carrier methods, some vf (voice frequency) pairs would be required for local circuits, supervisory pairs, etc. These vf pairs could be either incorporated together with the carrier quads or coaxial tubes in one cable in the form of one or more layers of audio pairs or, alternatively, separate vf cables could be provided

Carrier quads could be provided in either 40 pounds (lb)/mile or 25 lb/mile gage; vf pairs in either 40 lb/mile or 20 lb/mile gage. With cables laid adjacent to a 25-kv a-c electrified track adequate screening from induced voltages had been provided.

In accordance with railway requirements all telecommunication cables had to be placed in one concrete troughing laid along the track. Allowance for loading of all vf pairs and for appropriate cable accessories were included. Spacings of repeaters for different systems of working are shown in Table I.

It was assumed that reliable power supplies are available in all future repeater sites and therefore power feeding was not contemplated. For coaxial schemes allowance was made for suitable voltage regulators.

Method of Analysis

Altogether the study examined three carrier systems using two kinds of cable; the study allowed for two different gauges and for various quantities of vf and carrier pairs. Combinations of the above parameters required the inclusion of 28 different cables.

It is clear that all possible permutations of the above parameters would give a very large number of theoretically possible solutions. Hence a methodical step-by-step method was introduced for the analysis of the problem. Its logical basis can be formulated as follows. It is considered that the following parameters are at play and can be changed:

1. Number of cables, n
2. Size of cables, s
3. Additional facilities, f
4. Simplified techniques, t

Every one of these parameters is in turn dependent on at least three variables:

Price of copper, x
Price of lead, y
Other functions determining the cost of a system, z

If Q = total first cost of a system then

$$Q = F(n,s,f,t) \qquad (1)$$

and as each of the parameters in equation

Paper 58-905, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 19, 1958; made available for printing April 25, 1958.

ROMAN KRZYCZKOWSKI is with the Westinghouse Electric Corporation, Baltimore, Md.

Table I. Spacing of Repeaters

System	Method of Working	Conductor Gage, Lbs/Mile	Repeater Spacing, Miles	No. Required
12 channel carrier	2 wire	40 25	20 17	7 9
24-channel carrier	2 wire	40 25	12 9	13 17
Coaxial	Single 0.375-inch tube		6	27

Table II. Method of Analysis

Movement	Step	Carrier System	Cables Used	Combination
Starting point: what railways want	Evaluate the cost of original railways requirements (4 separate cables)	12	24 pr. carrier 40 lb 54 pr. trunk 40 lb 24 pr. trunk 40 lb 24 pr. trunk 40 lb	1 ↓
Diminish number of cables	Combine 24 pr. carrier and 54 pr. trunk	12	24/40c+52/40t 24/40t 24/40t	2 ↓
	Combine 24/40 and 24/40 trunks	12	24/40c+52/40t 54/40t	3 ↓
	Use *one* cable	12	24/40c+118/40t	4
Diminish size of cable	Use 20 lb for vf	12	24/40c+148/20t	5
	Use only 14 pr. for carrier circuits	24	14/40c+116/20t	6 ↓
	Use only one tube for carrier circuits	Coaxial	0.375-inch tube+120/20t	7 ↓
	Reduce number of vf pairs from 120 to 68	Coaxial	0.375-inch tube+68/20t	8
Give additional facilities	Separate carrier from audio	Coaxial	0.375-inch tube 54/20 trunk	9 ↓
	Give separate remote control cable	Coaxial	0.375-inch tube 54/20 trunk 6/20 trunk	10 ↓
	Increase number of vf pairs	Coaxial	0.375-inch+68/20t 6/20 trunk	11 ↓
	Use 40 lb gage for control cable	Coaxial	0.375-inch+68/20t 6/40 trunk	12
Simplify techniques	Substitute 24 channel for coaxial	24 channel	14/40c+52/20t 6/20 trunk	13 ↓
	Use vf for some medium length links, i.e., 40 lb gage	24 channel	14/40c+40/40t 6/20 trunk	14
Further increase facilities	Offer vf pairs in two cables	24 channel	14/40c+52/20t 24/20 trunk	15 ↓
	Offer some vf pairs in 40 lb gage	24 channel	14/40c+52/20t 24/40 trunk	16 ↓
	Give separate control cable	24 channel	14/40c+92/40t 6/20 trunk	17

c = carrier pairs
t = vf or trunk pairs

1 is a function of variables x, y, and z, then:

$$\begin{aligned} n &= F_1(x,y,z) \\ s &= F_2(x,y,z) \\ f &= F_3(x,y,z) \\ t &= F_4(x,y,z) \end{aligned} \qquad (2)$$

where $F_1, \ldots F_4$ are the respective functions.

The problem is to find the true minimums of equation 1. As the total differential from equation 1 is:

$$dQ = \frac{\partial F}{\partial n}dn + \frac{\partial F}{\partial s}ds + \frac{\partial F}{\partial f}df + \frac{\partial F}{\partial t}dt \qquad (3)$$

we can examine the problem methodically by keeping at a time all parameters but one constant, and varying the one under investigation. The influence of this variation on the total outlay is then observed. Equation 3 is thus the logical basis of the method employed.

As there are four parameters, four principal "movements" or variations in respective parameters, are examined:

1. Diminish number of cables—i.e. vary n
2. Diminish size of cables—i.e. vary s
3. Give additional facilities—vary f
4. Simplify techniques—vary t

Altogether 17 combinations are examined. In every case the layouts of the cables used are dictated by technical requirements of stranding.

The analysis starts from the original requirements suggested by the railway authorities:

1 carrier cable suitable for 12-channel working, 24 pairs/40 lb

1 long distance vf cable, 54 pairs/40 lb

1 short distance (local) vf cable, 24 pairs/40 lb

1 remote control cable, 24 pairs/40 lb

The total first cost of the preceding cables placed in concrete troughing, jointed, loaded, and brought into service, plus the cost of terminal and line equipments, gives the starting datum point. This is labelled combination 1.

Every movement is carried out in several steps. Thus the movement "Diminish number of cables" is done in four steps resulting in four combinations. The whole sequence of the analysis is shown on Table II which diagrammatically illustrates the adopted logical analysis. While it is realized that the "transportation" method of linear programming would give the same result, the way adopted here is preferred, as throughout the calculation a clear physical picture of result is apparent and nothing is left to the mechanism of tabulation. The position might have been different with a greater number of variables.

It is to be noted that carrier systems become the dependent function of the cable parameter. Thus the movement "Diminish size of cable" (in combination 6 and 7) requires the transformation of carrier techniques from 12 channels into 24 channels and coaxial respectively. A deliberate effort to consider the system as an independent parameter is made in combinations 13 and 14, which are the variations of t.

When it is required to find the influence of variables x, y, z on the total first cost, equation 1 can be expanded into:

$$\begin{aligned}
\frac{\partial Q}{\partial x} &= \left(\frac{\partial Q}{\partial n}\right)\left(\frac{\partial n}{\partial x}\right) + \left(\frac{\partial Q}{\partial s}\right)\left(\frac{\partial s}{\partial x}\right) + \left(\frac{\partial Q}{\partial f}\right)\left(\frac{\partial f}{\partial x}\right) + \left(\frac{\partial Q}{\partial t}\right)\left(\frac{\partial t}{\partial x}\right) \\
\frac{\partial Q}{\partial y} &= \left(\frac{\partial Q}{\partial n}\right)\left(\frac{\partial n}{\partial y}\right) + \left(\frac{\partial Q}{\partial s}\right)\left(\frac{\partial s}{\partial y}\right) + \left(\frac{\partial Q}{\partial f}\right)\left(\frac{\partial f}{\partial y}\right) + \left(\frac{\partial Q}{\partial t}\right)\left(\frac{\partial t}{\partial y}\right) \qquad (4) \\
\frac{\partial Q}{\partial z} &= \left(\frac{\partial Q}{\partial n}\right)\left(\frac{\partial n}{\partial z}\right) + \left(\frac{\partial Q}{\partial s}\right)\left(\frac{\partial s}{\partial z}\right) + \left(\frac{\partial Q}{\partial f}\right)\left(\frac{\partial f}{\partial z}\right) + \left(\frac{\partial Q}{\partial t}\right)\left(\frac{\partial t}{\partial z}\right)
\end{aligned}$$

Equation 4 was not directly used in the study but this formulation was at the back of investigation into risks involved in copper and lead price variations (variables x and y). It also served to show the lack of economic justification for using concrete troughing (variable z).

Results

Fig. 1 shows that all combinations examined offer respectively about 400 circuits, carrier and vf, when fully utilized. Starting capacities are near the 200 mark. Expansion possibilities of systems vary between 170% and 240% depending on

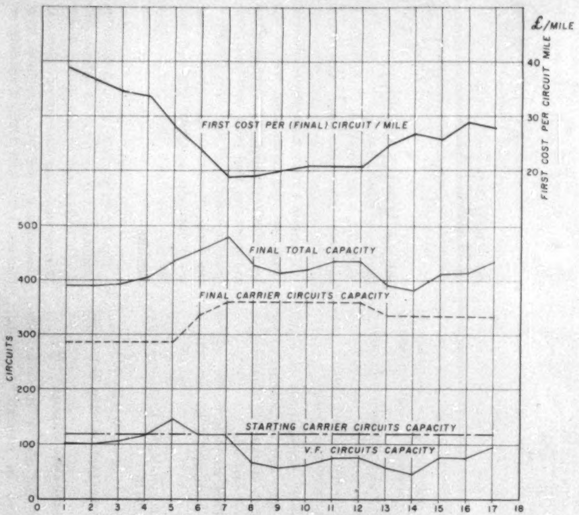


Fig. 1. Capacities and costs per circuit per mile

Table III. Circuit Capacities of Examined Systems

Combination	Total vf Circuits	Total Starting Capacity	Final Carrier Circuits Capacity	Final Total Capacity	Expansion Possibilities, %
1	102	222	288	390	175
2	100	220	288	388	175
3	106	226	288	394	175
4	118	238	288	406	170
5	148	268	288	436	165
6	116	236	336	452	190
7	120	240	360	480	200
8	68	188	360	428	230
9	54	174	360	414	240
10	60	180	360	420	230
11	74	194	360	434	220
12	74	194	360	434	220
13	58	178	336	394	220
14	46	166	336	382	230
15	76	196	336	412	175
16	76	196	336	412	175
17	98	218	336	434	200

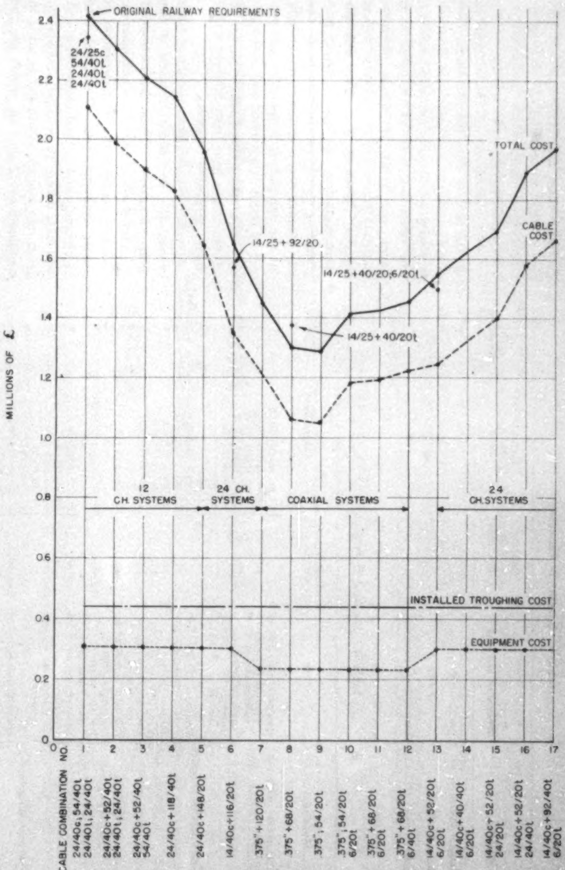


Fig. 2. Total costs

cable sizes. Coaxial solutions are found in that respect to be the most flexible. Detailed data on which Fig. 1 is based is given in Table III.

Fig. 2 shows the build-up of costs and to draw attention to their singular composition the first costs of two combinations (1 and 9) are drawn pictorially in Fig. 3. It is observed that by adopting coaxial carrier techniques the first cost of cables can be very appreciably reduced. The main difficulty in the further lowering of these costs is the outlay on troughing which remains fixed irrespective of the system adopted, and which in the least first cost solution, combination 9, is almost equivalent in magnitude to the cost of cables, while far exceeding the total outlay on equipment. Tables IV and V give the first costs of the combinations examined.

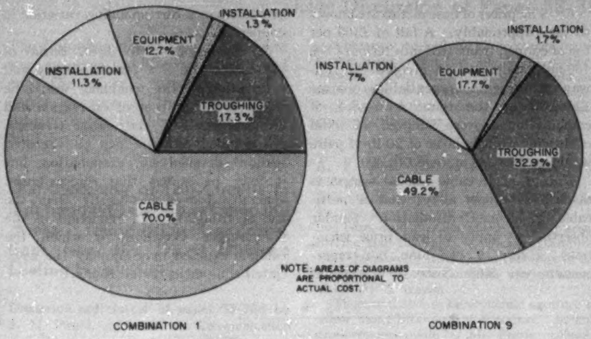


Fig. 3. Build-up of first costs

Table IV. First Costs of Systems Examined

Combination No.	Carrier System	Cables Used	Cost: Cables Per 1,000 Yards, £	Cost: Cables Per 158 Miles, £	Cost: Equipment, £	Total First Cost, £
1	12 channel	24/40c 54/40t 24/40t 24/40t	7,532	2,110,000	310,000	2,420,000
2	12 channel	24/40c+ 52/40t 24/40t 24/40t	7,133	1,995,000	310,000	2,305,000
3	12 channel	24/40c+ 52/40t 54/40t	6,788	1,900,000	310,000	2,210,000
4	12 channel	24/40c+ 118/40t	6,534	1,830,000	310,000	2,140,000
5	12 channel	24/40c+ 148/20t	5,813	1,650,000	310,000	1,960,000
6	24 channel	14/40c+116/20t	4,831	1,350,000	300,000	1,650,000
7	coaxial	0.375-inch+120/20t	4,340	1,215,000	235,000	1,450,000
8	coaxial	0.375-inch+68/20t	3,805	1,065,000	235,000	1,300,000
9	coaxial	0.375-inch tube 54/20t	3,758	1,050,000	235,000	1,285,000
10	coaxial	0.375-inch tube 54/20t 6/20t	4,227	1,180,000	235,000	1,415,000
11	coaxial	0.375-inch+68/20t 6/20t	4,274	1,195,000	235,000	1,430,000
12	coaxial	0.375-inch+68/20t 6/40t	4,398	1,225,000	235,000	1,460,000
13	24 channel	14/40C+52/20t 6/20t	4,497	1,260,000	300,000	1,560,000
14	24 channel	14/40c+40/40t 6/20t	4,753	1,330,000	300,000	1,630,000
15	24 channel	14/40c+52/20t 24/20t	4,984	1,400,000	300,000	1,700,000
16	24 channel	14/40c+52/20t 24/40t	5,690	1,590,000	300,000	1,890,000
17	24 channel	14/40c+92/40t 6/20t	5,960	1,670,000	300,000	1,970,000

Table V. First Costs of Systems Employing 25-Lb Carrier Conductors

Carrier System	Cables Used	Cost of: Cables Per 158 Miles, £	Cost of: Equipment, £	Total First Cost, £
12 channel	24/25c 59/40t 24/40t 24/40t	2,000,000	340,000	2,340,000
24 channel	14/25c+92/20t	1,240,000	330,000	1,570,000
24 channel	14/25c+40/20t	1,045,000	330,000	1,375,000
24 channel	14/25c+40/20t 6/20t	1,170,000	330,000	1,500,000

Conclusions

1. The least cost solutions are obtained by using coaxial carrier techniques, the least number of cables, and an absolute minimum of vf pairs.

2. For the three carrier techniques under comparison the first costs of terminal and line equipment are broadly similar, and from other studies it is known that this is broadly true within limits for lengths of line 50–500 miles, and for circuit capacities 100–1,000 telephone channels. Expenditure on equipment forms about one fifth of the total expenditure and hence savings on equipment cannot have a decisive influence on the over-all outlay.

3. The cables under consideration offer approximately either 50 or 100 vf pairs to be used for local and control circuits. Some pairs will be definitely required and one layer, that is approximately 50 pairs is the smallest practicable quantity to be incorporated in a composite cable. The cost of the next 50 pairs increases the cost of any system by approximately £1,000 per mile.* Thus from the first cost point of view there is a strong case for reducing the number of vf pairs to one layer only, and engineering most of the connections as carrier circuits even for relatively short distances.

4. Economies available on account of diminishing the number of cables are in the order of £2,000 per mile when one composite cable only is employed instead of four separate ones.

5. The first cost of the convenience of having 40 lb/mile vf conductors instead of 20 lb conductors is broadly £600 per mile for the first layer (about 50 pairs)

---

* A pound (£) is presently worth $2.80.

and £1,300 per mile for the next 50 pairs. These figures contain the additional expenditure on lead sheathing and armouring of cables of larger diameter.

The cost advantage of 25 lb gauge of carrier conductors is the greatest when 24/40 carrier conductors are replaced by smaller gage in conjunction with a 12-channel system. The cost advantage is then in the order of £600 per mile. Fig. 2 shows however that no significant economies are obtained in the region of the least cost solutions. Twenty-five lb/mile conductors are not competitive enough for coaxial tubes.

6. The prices of base metals are known to vary appreciably. A fall of £100 per ton of copper, translated into terms of the 160-mile route, would involve the railway authorities in a potential loss of some £150,000 on the route concerned, if combination 1 were adopted. Coaxial cable with a single layer of 20 lb vf pairs would entail a loss of some £40,000.

Thus a single cable coaxial system is much less "copper risky" than a multicable 12-channel installation. Similar observations apply to lead price variations. Every £ per tone has repercussions on cable costs several times greater than a corresponding variation in copper price.

7. By resorting to direct burial of the cables, the cost for the supply and installation of the troughing would be replaced by the cost of excavation and reinstatement of the cable trenches. No detailed estimate can be ventured without detailed site examination, but it may be stated that direct burial generally proves less costly than the use of troughing and reduction of the temperature variation to which the cable is exposed may simplify the associated equipment requirements.

◆

Discussion

G. Rezos (The Pacific Telephone and Telegraph Company, San Francisco, Calif.): This is a very interesting and unusual method for deciding which is the best plan to use for a new cable carrier system.

There are a few points which need clarification:

1. The time interval between the "starting" capacity and the "final total" capacity should be specified. If it is a slow growing system, we can justify a very expensive ultimate layout, if we can defer high cost items for a long time. Then we can save large amounts in interest, carrying charges, depreciation reserves, etc.

For a fast growing system, where we must spend our money over a few years, low overall cost becomes all-important. In other words, *when* we spend the money is just as important as *how much*, and the when may become controlling. Early expenditures have much more effect than later ones. Also, to compare several plans we should have equal starting capacity. Hence it seems risky to compare combination 5 (starting with 268 circuits) with combination 9 (174 circuits) even though their final total capacity is about the same (436 and 414). We would have to spend a lot of money earlier on combination 9 to make it match combination 5.

2. Table IV shows equipment costs for all combinations to be roughly the same. While this may be true when any combination reaches its final total capacity, it does not seem reasonable for the initial costs. The "get started" costs for a small number of coaxial circuits is generally very much higher than that for a few 12-channel cable carriers.

3. Intangible factors must be considered. In the coaxial plans, the failure of a single repeater element could disable all through circuits; with several 12-channel cable carriers a complete failure from equipment troubles is much less likely. On the other hand, a complete break in the cable itself would affect coaxial and cable carriers equally.

4. The decision between direct burial of the cables or troughing could have a major effect on the choice of lowest cost combinations. A direct buried cable must be much larger because it must last many more years than cables placed in conduit. In conduit, cables can be added at economic intervals; with direct burial the hazards of burying a second cable near the first may be out of question. The greater first cost of the larger buried cable may completely override the effects of other variables.

Roman Krzyczkowski: Mr. Rezos is quite right in pointing out the danger of limiting economic studies to first cost considerations only. "*When* we spend the money is just as important as *how much*" is one of the best formulations of the problem I have ever seen.

I deal rather fully with this subject in another paper[1] and, incidentally, I point out there that inflation is yet another factor which necessitates taking *time* into consideration.

The present paper describes only a fragment of a complete economic study. This fragment is concerned with the determination of the least first cost—as the title of the paper implies—and not with final decision: which is the best plan to use for a new cable carrier system. Other parts of the study, apart from calculations of (time) capitalized costs, dealt also with intangibles, which, as Mr. Rezos rightly points out, are of decisive importance.

I agree that the choice between direct burial of the cables and troughing makes a fundamental difference in first costs. The decision to put cables in troughing was arrived at from consideration of intangibles, such as safety, reliability, and maintenance.

European and American prices are different and consequently it is misleading to make direct dollar comparisons. For that reason original figures are left in pounds sterling and pounds weight and are not expressed in dollars and wire gages. Also, it is the primary purpose of the paper to describe a method of determining the least cost. According to my experience, as soon as the "get started" capacity is in the order of one full International Consultative Committee on Telegraphy (CCIT) supergroup, initial costs of a coaxial and of 4-wire multicore cable installations break even. This appears to be confirmed by reference 2.

References

1. INTEREST AND INFLATION IN THE ECONOMIC PLANNING OF TELECOMMUNICATIONS SYSTEMS, **R. Krzyczkowski.** *ATE Journal*, Liverpool, England, vol. 13, 1957, p. 298.

2. PROGRAMME GÉNÉRAL D'INTERCONNEXION TÉLÉPHONIQUE EN EUROPE. International Consultative Committee on Telephony (CCIF), Montreux, 1946, App. 16.

Late Discussion of Paper 58-386

The following discussion and closure were received too late to be published with the paper; they are included herein because of the similarity of subject matter to the preceding paper 58-905.

Dial Switching of Toll Circuits in Independent Telephone Company Areas: Transmission Considerations from the Direct Distance Dialing Standpoint

Discussion and closure of paper 58-386 by J. N. Petrie, published in Communication and Electronics, September 1958, pages 481–85.

E. E. Nelson and **George Rezos** (The Pacific Telephone and Telegraph Company, San Francisco, Calif.): In reference to the paper by J. N. Petrie, we would like to discuss the following considerations.

1. The paper assumes the continued use of CDO (community dial office) tandem points. It proposes "tail-eating" switching at such points (class *5-A*). We are basing our plans on getting rid of tandem CDO operation as soon as possible to avoid transmission problems, as well as dialing problems such as limited stop-go features in outgoing crossbar tandem senders. The change from tandem CDO operation would not be done by itself; it would be combined with some major change to reduce costs.

2. Balancing problems are also involved in switching the line and network leads. To avoid unbalances, the line and net leads should have the same resistance and capacity. The net leads should follow the same path through the office as the line, including the path through the trunk circuits. Quadded switchboard cable will probably be necessary, using one side for line and the other side for net. Repeating coils in the trunk circuit should be matched in the net path. The physical location of the balancing equipment could become troublesome.

3. In normal circuit order work involving jumper cross connections, tip and ring turnovers will have to be watched closely to combine the power in the tandem hybrid coils.

4. The 4-db (decibel) loss of the hybrid will be obtained in the terminating condition (3-db division of power + 1-db copper, core, and wiring loss). The 1-db loss of the hybrid in the through condition should be compensated by line and net pads in the switching train. The pads, relays, and control circuits could run into quite a bit of space and complexity.

5. Extra contacts are required on all switches in the toll train. Class *4* offices with toll switchboards would require 4-wire switching. In the class *5 A* offices a large part of the 4-wire switches would be connecting a compromise network to the hybrid.

6. The "tail-eating" scheme was attractive when first proposed many years ago because we had to interconnect many 2-wire trunks to a few 4-wire (or carrier) trunks. Now, with most toll circuits on carrier, and the increasing use of carrier even for very short hauls, the advantages gradually disappear.

A. F. Rose (American Telephone and Telegraph Company, New York, N. Y.): Tail-eating switching as described in Mr. Petrie's paper is not new in principle. It has been used in Australia and other areas for the last 15 years. I assume, therefore, that the novelty involved in this case is the Automatic Electric Company version known as the *FW-1* system.

There will always be economic pressure to allow more links in a toll connection, and the proposal made by Mr. Petrie probably involves less transmission sacrifice than would be likely if 2-wire switching of the ordinary type were used. However, I feel he has underestimated the transmission penalty, as Bell System experience shows that transmission variations added by each link are presently considerably exceeding the 0.4 db anticipated when the via net loss factors were established. Also, it should be borne in mind that an additional link at one end of the connection may also involve one at the other end.

I am glad to see that Mr. Petrie appreciates the problems of PBX (private branch exchange) transmission and compares his proposal for the *5 B* office to the use of a PBX trunk as an extra link in the toll chain. This agrees with our thinking but we feel that as direct distance dialing is established into and out of PBX's we may have to treat them as end offices in some cases and home them on suitable higher ranked offices. Some large PBX's in the Bell System are already being handled in this manner, and this serves to improve rather than degrade transmission to these important customers.

In some of Mr. Petrie's illustrations, "return loss" objectives are indicated. It is not clear at what point in the switching office these figures apply. In Bell System usage we refer to return loss measured at the switches. We find that the average return loss of loops against a compromise network is 11 db with a sigma of 3. If a measurement were made at the hybrid through a 5-db pad this would be 11 db + 2 × 5, or 21 db. On the other hand, if the intention is to measure at the switch points, 20 db in a class 5-office is somewhat better than our experience would indicate to be realizable.

Finally, it seems worthwhile to point out that spending the extra capital to provide this type of switching does not completely eliminate the maintenance attention needed to keep office balance up to its original value. Unbalance can occur even under the tail-eating scheme, and reversal of tip and ring may have more serious consequences.

In bringing these problems to the attention of the independent companies, I believe Mr. Petrie has made a considerable contribution. However, I hope that they will understand that adding links to a toll connection adds transmission loss even with the best switching system and should be resisted unless an economic study shows substantial savings. Where we have studied similar cases, the cost advantage has favored direct circuits from the *5B* office to the toll center instead of special switching equipment. With the continuing downward trend in the cost of carrier facilities this is even more likely to be true in the future.

J. N. Petrie: The six considerations contained in the discussion by Mr. Nelson and Mr. Rezos are all quite pertinent. My comments will take up each in turn.

1. From the Bell System point of view there can be no argument with the philosophy expressed by these gentlemen; they seem to have missed the point, however, of some of my remarks in the opening paragraphs that: "The problem of complying with transmission requirements of the general toll switching plan, and also of fitting into classifications established for them, is particularly difficult for many independent telephone companies." Why this is so is discussed in the next two paragraphs following the one from which the quotation was taken. To "legitimize" this situation would require the Bell System to upgrade many class *4* (toll center) offices to class *3* (primary center) offices.

2. All of the problems mentioned under this heading have been recognized and have been solved, essentially as suggested, in the *FW-1* switching scheme which is described in several of the references included in the original paper.

3. This is not a jumper or cross-connection problem as the 180 degree phase reversal is handled automatically by the switching equipment.

4. There is no argument with the several points raised. Nevertheless, I am reliably informed that the space and circuit complexity problems are no greater than with other 4-wire switching schemes.

5. Four-wire switching at class *5A* offices, on the basis suggested in the paper, does not necessarily require 4-wire switching at the class *4* (toll center) office. Figs. 5 and 6 show the two alternatives. It is true that the switches at a *5A* office would be connecting a compromise network part of the time. This is one of the advantages of the system proposed since it permits each switch to be made under the most favorable conditions and does not commit a whole office to be built-out (at great cost and with recurring high maintenance) to a compromise condition. The implication seems to be that the 4-wire switches would be too costly. On this point I am informed that the extra cost for 4-wire switches versus 2-wire switches is of the order of 10%, which, it would appear, would be very reasonable for the gains and benefits derived.

6. Whether the advantages of the tail-eating connection are gradually disappearing will depend on the switching philosophy which a company may adopt. Certainly, for the smaller trunk groups, typical of most independent companies, there are advantages in being able to terminate or to tandem switch indiscriminately all carrier circuits, regardless of make, different transmission levels, and circuit impedances at the 4-wire points. The tail-eating techniques of switching do offer this advantage.

In reply to the very pertinent comments by Mr. Rose, I would like to point out that the pressure for the switching arrangement proposed is not primarily economic but rather of an administrative nature. This results largely from the need of many independent companies to operate a "community of interest" telephone network with toll center functions concentrated at a class *5A* office. The proposed arrangement recognizes this situation and suggests a means of meeting the established toll transmission requirements in spite of the tandem switch added at the *5A* office.

The comments with respect to the return losses shown are entirely valid. It is, however, a fact that the insertion of pads, as indicated, was only partly prompted by return loss considerations; in fact, the values chosen were primarily dictated by gain switching requirements. The pads do have the effect of improving the return loss at the hybrid coil terminals, thus making gain switching possible.

It is recognized that the successful operation of the tail-eating switching scheme presupposes symmetry of circuits through an office but this is not in the same sense that office balance is a problem at 2-wire intertoll switching points. The reversal of the tip and ring conductors is no problem since the switches introduce the necessary 180-degree phase reversal. An accidental reversal introduces such effective transmission blockage that it should be no problem to locate.

There can be no agrument with Mr. Rose's statement that direct circuits from the end office to the toll center (class *4*) may have cost advantages. As stated earlier, the switching scheme under discussion is being proposed primarily to solve an administrative problem.

A Triple-Action Positioning Controller Using a Magnetic Modulator

C. V. BULLEN
NONMEMBER AIEE

THIS PAPER describes a process control of the triple-action positioning type; i.e., the control circuits provide variable proportioning, automatic reset, and rate actions to position a valve or other modulating device. The input or error signal is derived from a recording instrument of the null-balance type, which can be a potentiometer recorder operating from a thermocouple, a self-balancing resistance thermometer, or a pressure recorder operating from a strain-gage transducer.

The novel feature of this control circuit is the use of a magnetic modulator, whose purposes are

1. To add the error signal and the various feedback signals.
2. To convert the sum into an a-c signal which can be amplified readily to control the valve position.

The magnetic modulator assembly (Fig. 1) consists of a pair of transformers each having five windings. The core material is a high nickel alloy selected for a rapid change in permeability with a change in saturation. Corresponding windings on the two cores are connected in series. One pair of windings is connected to a 60-cycle-per-second source to provide excitation. A second pair of windings provides the output voltage. These windings are connected in series opposition so that, with no current in the control windings, the voltages from the two windings are equal and opposite and the net output from the modulator is zero. A third pair of windings carries a direct current which biases the cores to a steep part of the permeability curve. The fourth and fifth pairs of windings are control windings. Each pair of control windings is polarized so that a current in it adds to the bias mmf (magnetomotive force) in one core and subtracts from it in the other. Thus, the permeability of the core with the lower net mmf is higher and the voltage induced in its output winding predominates over the voltage in the output winding of the core which has the greater net mmf.

The magnetic modulator assembly is shown in Fig. 2. A somewhat simplified diagram of the complete control is shown in Fig. 3.

The error signal is derived from an auxiliary slide wire in the null-balancing instrument. The drum carrying the resistance winding is mounted on the main shaft of the recorder, while the slider position is adjustable for selection of control point. The slider mounting arrangement is such that the slider contacts the center of the resistance winding when the recorder registers the desired temperature. Control winding *1* of the magnetic modulator is connected from this slider to the center connection of two equal resistors connected in series across the same 8-volt source that energizes the auxiliary slide wire. Thus, when the recorded temperature is at the desired control point, the current through this control winding is zero. When the recorded temperature departs from the control point setting, an error current, varying in polarity and magnitude with the direction and degree of departure, flows in control winding *1*. This causes an a-c signal to appear at the output of the magnetic modulator. This signal is amplified by a 2-stage voltage amplifier and applied to the paralleled grids of a phase detector. The plates of the phase detector, with relays in series, are connected to opposite ends of a transformer secondary so that the supply voltages for the plates are 180 degrees apart in phase. When a signal is applied to the grids it will be in phase with the plate supply voltage of one of the triodes and a current will flow in the corresponding relay. The relay contacts control the operation of the motor which positions the valve or other modulating device.

The combination of magnetic modulator, voltage amplifier, and phase detector thus causes the valve to open or close whenever the recorder indication deviates below or above the desired control point.

The motor which drives the valve also drives a slide wire which is energized by the 8-volt source. The slider voltage, which indicates the valve position, is applied to one grid of a balanced cathode follower through a series capacitor *C1*. The output voltage of the cathode follower is connected to control winding *2* of the magnetic modulator. Normally, when the recorded temperature has been at the desired value for some time, both grids and, therefore, both cathodes are at the same potential and no current flows in control winding *2*. A balance adjustment is provided at the ground end of the cathode resistors to assure this condition regardless of inequalities in the two sections of the tube.

If the controlled temperature drops, an error current is caused to flow in control winding *1* and the valve motor is caused to run in the opening direction. The potential of the slider of the valve slide wire rises toward ground and the change in voltage is applied through *C1* to the left grid of the cathode follower. The potential of the left cathode rises by an almost equal amount and thus causes a current to flow in control winding *2*. As the valve continues to open, this current increases until it eventually equals that in control winding *1*. Since the mmf due to currents in the two control windings oppose each other, as soon as the current in winding *2* equals that in winding *1*, the output of the magnetic modulator drops to zero and the valve motor stops. The amount of valve motion necessary to accomplish this rebalance after a given deviation from the desired temperature can be varied by adjustment of a rheostat in series with control winding *1*. This rheostat is calibrated in per cent proportional band which is the temperature deviation in per cent of full recorder scale divided by the resulting valve motion in per cent of full travel, the quotient also being expressed in percentage. In

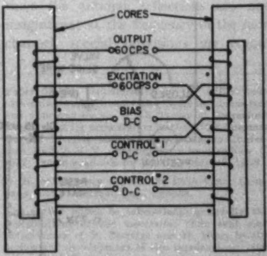


Fig. 1. Magnetic modulator

Paper 58-850, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted December 31, 1957; made available for printing April 14, 1958.

C. V. Bullen is with the Barber-Colman Company, Rockford, Ill.

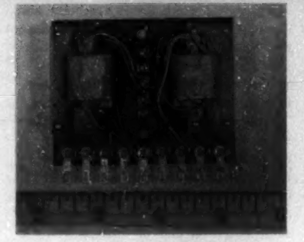


Fig. 2. Magnetic modulator

this control the proportional band may be adjusted through a range of 2% to 200%.

The type of control described in the foregoing is known as proportional control and is characterized by a corrective action proportional to the error. It allows control without hunting but usually results in the controlled condition being maintained at some value other than desired. This is the control point shift associated with straight proportional control. To avoid this shift it is necessary to introduce resetting, also known as floating or integral control. In this type of control, the amount of correction is proportional to the time integral of the error.

Capacitor *C1* starts to charge to the potential of the valve slider and, as it charges, the potential of the left grid of the cathode follower falls back toward zero. The current in control winding *2* drops, upsetting the balance previously obtained in the modulator. This results in a further small corrective motion of the valve. As capacitor *C1* continues to charge, periodically unbalancing the modulator, the valve continues to inch along, always in the same direction as the initial motion. The average rate of valve motion is determined by the time constant of *C1* and the adjustable resistor *R1*. The reset rate is expressed in repeats per minute, which is the average valve rate in degrees per minute divided by the initial valve motion (due to proportional action) in degrees, assuming a constant error. In this control, reset rate is adjustable between 0.06 and 100 repeats per minute.

The resetting action continues until the error has been reduced to zero and *C1* has fully charged to the slider potential, reducing the potential of the left grid of the cathode follower to zero.

To minimize overshooting, it is desirable to introduce a third type of control in which valve motion is a function of rate of change of error. This introduces a

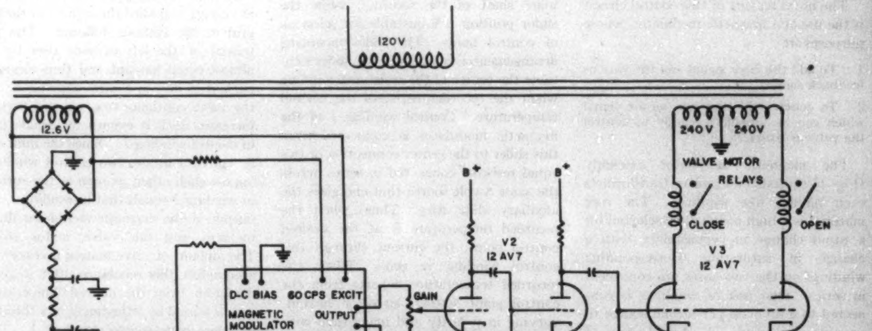

120V
12.6V
-8V
50
50
RECORDER SLIDE WIRE
HIGH
LOW
D-C BIAS
60 CPS EXCIT
MAGNETIC MODULATOR
OUTPUT
GAIN
CONTROL 1
CONTROL 2
PROP BAND
B+
V2
12 AV7
B+
240 V
240V
VALVE MOTOR
RELAYS
CLOSE
OPEN
V3
12 AV7
-8V
VALVE SLIDE WIRE
CLOSED
OPEN
V
POSITION
C1- 15 MFD
RESET RATE
0-70MΩ
R1
B+
V1
12 AV7
C2- 15 MFD
RATE TIME
R2
0-6MΩ
BALANCE

Fig. 3. Position control circuit

correcting factor which adds to, or subtracts from, the effect of the error magnitude. The rate correction is introduced at the right grid of the cathode follower by the differentiating network *C2*, *R2*. The effect of this connection is to apply a potential at the right grid which is a function of the rate of change of error voltage. For example, if the error is positive and rising, a negative potential is applied to the right grid. This makes it necessary for the valve to move further in the closing direction to rebalance the modulator than it would if the velocity component were absent, and the overshooting tendency is reduced.

Rate effect is measured by per cent of valve movement per per-cent per minute of recorder rate. The units work out to be minutes. In this control rate times of up to 5.5 minutes can be obtained by adjustment of *R2*.

The reset and rate circuits are essentially the same; in each a varying voltage is connected to a resistor and capacitor in series, with an output voltage taken from across the resistor. However, the actions of the two circuits are quite different because of differences in time intervals and time constants. In a circuit of this configuration:

$E_0 = KRC(1-\epsilon^{-t/RC})$ when $E_1 = Kt$

where

E_1 = input voltage (across resistor and capacitor in series)

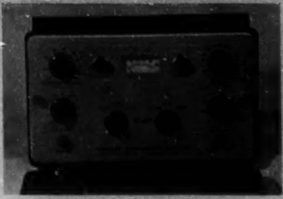


Fig. 4. Panel of control unit

E_0 = output voltage (across resistor)

K = rate of change of input voltage, volts per second

R = resistance, ohms

C = capacitance, farads

t = time after input voltage starts to change, seconds

$\epsilon = 2.7183$

In the reset circuit, a typical value for *RC* is 800 seconds, whereas the valve motor may run 10 seconds or less to rebalance the circuit. The ratio t/RC is thus very small and under this condition $E_0 = Kt$ very nearly. Therefore, the output voltage is practically the same as the input voltage.

In the rate circuit, a typical value of *RC* is 5 seconds. The input voltage may continue to rise for minutes. Where t is several time constants $E_0 = KRC$ very nearly and, therefore, the output voltage is a linear function of the rate of change of input voltage.

Fig. 5. Control unit mounted on recorder

A valve position meter consisting of a voltmeter connected from closed end to slider of the valve slide wire is provided. Auxiliary switches allow the valve to be positioned manually when desired.

Fig. 4 shows the panel of the control unit. It is ordinarily mounted below the recorder, as shown in Fig. 5. The door of the control has an opening which exposes only the controls necessary for starting and shutting down the system. When the control door is closed and the recorder door locked, none of the critical adjustments are accessible to unauthorized persons.

A Quadruplexer Allowing the Simultaneous Transmission of Two Complete Television Stations Using a Common Antenna

G. B. MacKIMMIE
NONMEMBER AIEE

IN THE SPRING of 1956, the need arose in Quebec City to add a new English-language television station in order to supplement the existing service of CFCM-TV, in operation on channel four, and serving primarily the predominant French-speaking population.

The transmitting facilities of CFCM-TV consisted of a 5-kilowatt transmitter, and a 3-bay superturnstile antenna mounted on a 500-foot self-supporting tower. In the interest of greatest efficiency and economy, the owners were anxious to make maximum use of existing facilities in planning the new station, which later became known as CKMI-TV and which operates on channel five. A plan was therefore evolved which included use by the new station of the existing antenna and feed system, with no alteration, and without affecting in any way operation on channel four.

In order to accomplish this result, a "quadruplexer" was designed and constructed. The quadruplexer provides a means of interconnecting the four transmitters, two aural and two visual, to the antenna feed system with no interaction among the transmitters. Furthermore, since the antenna reflections are not insignificant at the frequency of the new channel, special provisions are incor-

---

Paper 58-904, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y. Manuscript submitted Nov. 6, 1957; made available for printing May 12, 1958.

G. B. MacKimmie is with the RCA-Victor Company, Ltd., Montreal, Que., Canada.

The author wishes to acknowledge gratefully the contributions of his associates. The skill and perseverence of B. Berridge and H. Hings in the design and development of this quadruplexer contributed vitally to its success. In addition, the kind co-operation of the owners and engineers of CFCM-TV and CKMI-TV during installation and final testing is greatly appreciated.

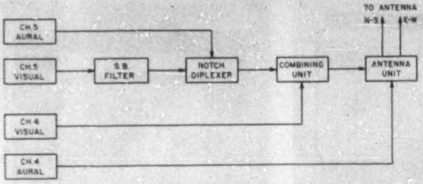


Fig. 1. System block diagram

porated within the quadruplexer to absorb channel five antenna reflections and thus to prevent the radiation of a ghost signal. It should be noted in passing, that, though channels four and five are numerically adjacent, there is actually a spacing of 4 megacycles (mc) between them, and therefore assignment of both in a city is permissible. Channel four covers a frequency range from 66 to 72 mc and channel five from 76 to 82 mc.

It is the purpose of this paper to describe the quadruplexer, which has now been installed successfully and has been in commercial operation since March 1957.

General Description

The basic elements of the quadruplexer are shown in the block diagram, Fig. 1. The channel five visual signal, after passing through a vestigial sideband filter, is combined with the channel five aural signal in the notch diplexer, the first element of the quadruplexer. The single output of the notch diplexer feeds one input of the combining unit, which also receives at another input the channel four visual signal. The output of the combining unit is delivered to the antenna unit, which also receives the channel four aural signal, and delivers the complete dual-channel signal to the antenna transmission lines. Since the antenna is of the turnstile type, all north-south radiating elements are brought together and fed from one line, while all east-west elements are similarly fed from a second line. To obtain nondirectional radiation from the antenna, the east-west line is made 90 electrical degrees longer or shorter than the north-south line in the usual way.

The antenna unit incorporates the channel five reflection absorption device mentioned previously. If this feature had not been provided, there would have been excessive ghost radiation on channel five, since the measured voltage standing wave ratio due to the antenna termination on this channel is in excess of 1.2. On channel four there is, of course, no reflection problem, since the antenna was originally designed and erected for operation on channel four.

Circuit Description

Having outlined the composition of the quadruplexer in its three essential parts, the notch diplexer, the combining unit, and the antenna unit, there will follow a brief description of each of these units. Before proceeding, however, a review of the properties of the bridge-balun, or more simply bridge, will assist in describing the circuits to follow.

The Bridge-Balun or Bridge

Fig. 2 is a pictorial outline of the bridge which is used in a number of places in the quadruplexer circuits. The bridge is essentially a 4-terminal hybrid and may take on any one of a variety of forms. The form shown, however, is particularly adapted to high power-television usage.

Referring to Fig. 2, and assuming the the output terminals to be terminated in loads of equal impedance, the application of a voltage at the end input will result in equal and out-of-phase load currents. No signal will appear at the side input. Similarly, a voltage applied at the side input will result in equal and in-phase load currents, and no signal at the end terminal. Voltages may be applied at both end and side inputs simultaneously, resulting in a vector addition of currents in one load and a vector subtraction of currents in the other load. At the same time, each input will be "blind" to the other input, and in the ideal case there is infinite isolation between inputs.

The bridge can operate also in reverse. Equal out-of-phase voltages applied at the output terminals will result in current flow only through a load connected at the end terminal, while equal in-phase voltage applied at the output terminals will result in a current flow only through a load connected to the side terminal.

The foregoing remarks, except those in the paragraph immediately preceding, apply when the output terminals are terminated in identical loads. Another case of special interest arises when the output terminals are terminated in transmission lines, differing in length by 90 electrical degrees, which are in turn terminated in loads of identical impedance. In this instance, when the loads are chosen so as to equal the characteristic impedance of the transmission lines, behavior of the bridge is as noted in the foregoing, except that due to the additional transmission line length in one load branch, the load currents are equal but in phase quadrature. This will be the case for input signals fed into either the end or side terminals. If the two loads remain identical, but do not equal the transmission line characteristic impedance, the equal amplitude and quadrature relation of load currents no longer

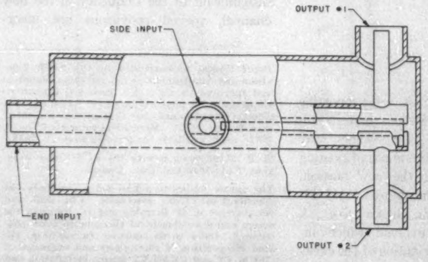


Fig. 2. Bridge-balun

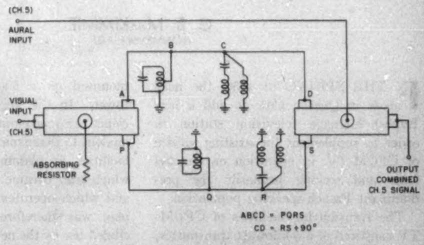
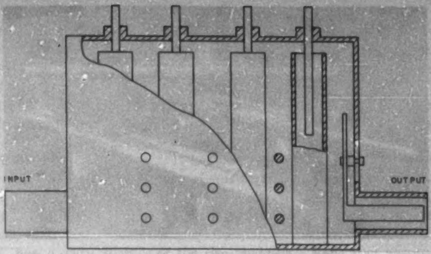


Fig. 3. Notch diplexer schematic

Fig. 4. Combining unit

Fig. 6 (right). Four-stub bandpass filter



holds, and in addition, the two inputs do not remain perfectly isolated. If, however, when feeding a signal to the end terminal, a properly selected load impedance is connected to the side terminal, it is possible to restore the original load current conditions. In addition, the input impedance at the end terminal will have the same value as for the condition of matched transmission lines. Upon close examination of this property of the bridge, one finds that the reflected signals returned along the transmission lines from the mismatched loads appear at the bridge output terminals in-phase, and accordingly feed the side terminal where they may, if desired, be absorbed. These same desirable results could be achieved while feeding the side input, if an absorbing load were connected at the end input.

Advantage is taken of these bridge properties to give the following important features:

1. To provide a constant resistive termination at filter inputs over stop-bands as well as over pass-bands.

2. To provide absorption of the channel five antenna reflection which, if radiated, would result in an objectionable ghost.

The internal construction of the bridge is so arranged, and dimensions are chosen so that 51.5-ohm terminations at the two output terminals result in input impedances at either end or side input terminals of 51.5 ohms over a limited band of frequencies. A particularly broad band input impedance may be obtained at the end terminal by suitable construction of the innermost conductor which may have two or more different diameters throughout its length.

Having summarized the essential characteristics of the bridge element, the first unit of the quadruplexer, the notch diplexer, may now be described.

Notch Diplexer

Combining of the channel five visual and aural signals is accomplished in the notch diplexer. Fig. 3 is a schematic diagram of this unit. All interconnections, including *ABCD* and *PQRS*, actually are made by means of coaxial transmission lines. The reactive elements at *B*, *C*, *Q*, and *R* are connected in shunt across the transmission lines at the points indicated. These elements are shown in the diagram as comprising simple inductive and capacitive components. Actually, they take the form of coaxial stub assemblies having properties similar to those indicated in the schematic representation. The reactive element at *C* consists of a simple series tuned circuit, sharply resonant at aural carrier frequency. The resulting capacitive reactance at the lower frequency of the visual carrier is antiresonated by means of the separate shunt inductor. The reactance element at *R* is identical with that at *C*.

At *B* and *Q*, identical circuits, tuned to antiresonance at visual carrier frequency are used.

The aural signal is delivered to the side terminal of the output bridge, resulting in equal in-phase voltages at *D* and *S*. As the signals proceed to the left, however, the shunt reactance elements at *C* and *R* are encountered. Since these are short circuits for the aural carrier frequency, reflected signals of 100% amplitude are returned towards the right, and to bridge terminals *D* and *S*. Because the electrical distance *DC* is 90 degrees longer than SR, the signals returning at *D* and *S* are now out-of-phase, and therefore appear at the output end terminal of the right hand bridge as desired. If the short circuits at *C* and *R* are not quite perfect, a small portion of the aural signal will continue

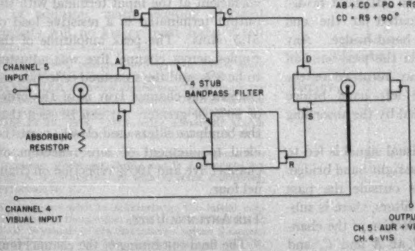


Fig. 5. Combining unit schematic

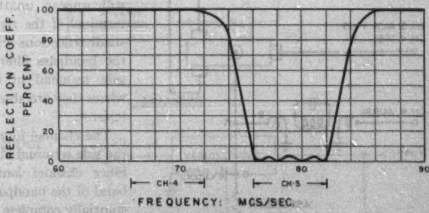


Fig. 7. Input reflection coefficient versus frequency for single 4-stub bandpass filter

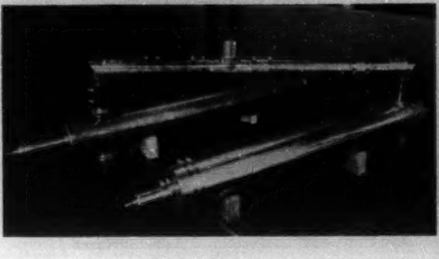


Fig. 8. Antenna unit

Fig. 10. Bridge

towards the left, and result in equal in-phase voltages at A and P. These voltages in turn direct a signal to the side terminal of the input bridge, where an absorbing resistor is used. No aural signal can occur at the end terminal of the left hand bridge.

The visual signal is delivered to the end input of the left hand bridge, so that equal out-of-phase voltages appear at A and P. At visual carrier frequency, all reactance elements are antiresonant, and therefore have no shunting effect. Equal and out-of-phase voltages therefore appear at D and S, causing the visual signal to be delivered to the end terminal of the right hand bridge as desired.

It is necessary, of course, that the diplexer pass without attenuation frequencies other than the visual carrier alone. It can be shown that the power lost in transmission of signals near visual carrier is equal to the square of the reflection coefficient on lines AB and PQ. Also, any such reflected energy will be absorbed in the absorbing resistor, and will not appear at the visual input terminal. The function of antiresonant reactive elements at B and Q is to reduce the reflection coefficient on lines AB and PQ in the portion of the visual spectrum above the color subcarrier. The spacing BC and slope of the susceptance characteristic of the element connected at B is chosen for maximum improvement. The over-all response characteristic is thus made flat within a few tenths of a decibel from 1 mc below to 4 mc above visual carrier frequency, and down less than 1 decibel at 4.2 mc above visual carrier frequency.

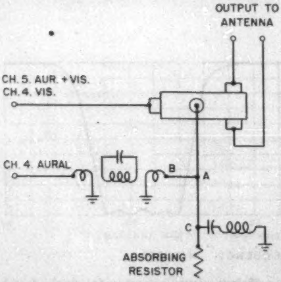


Fig. 9. Antenna unit schematic

Additionally, the notch diplexer circuit establishes a constant input termination at the visual input for signals of all relevant frequencies, regardless of whether or not attenuation takes place within the diplexer.

The Combining Unit

The channel four visual signal is added to the complete channel five signal in the combining unit; see Fig. 4. Referring to Fig. 5, it will be seen that the circuitry resembles very closely that of the notch diplexer. The principle difference is in the nature of the elements inserted in the transmission lines interconnecting the bridges. The two identical bandpass filters, BC and QR, are designed to have a flat response and very small reflection coefficient at their inputs over all frequencies of channel five. Therefore, because of the symmetrical nature of the circuit, it is evident that the complete channel five signal applied at the end terminal of the left hand bridge will appear unattenuated at the end terminal of the right hand bridge. Any small reflections within the pass band of the bandpass filters are returned to the side terminal of the left hand bridge where they are absorbed by the absorbing resistor.

The channel four visual signal is fed to the side terminal of the right hand bridge. Since channel four is outside the pass band of the bandpass filters, there is substantially complete reflection of the channel four visual signal at R and C, and therefore this signal is delivered with little attenuation to the end terminal of the right hand bridge along with the complete channel five signal. Small amounts of the channel four signal which may leak past the bandpass filters are absorbed by the absorbing resistor, but do not arrive at the channel five input terminal. Special measures were taken at the channel four visual input terminal of the combining unit to provide a broad band impedance match.

This input normally tends to be restricted in bandwidth because of the use of a side input of a bridge, and also because the electrical difference in the lengths CD and RS cannot remain exactly 90 degrees for all frequencies occupied by the visual signal. This difficulty was successfully overcome, however, by the use of a carefully designed half-wave broadbanding transformer in series with the channel four visual input.

Fig. 6 is a pictorial sketch of the bandpass filters used in the combining unit. These filters consist of four resonant stubs, mutually coupled together, and inductively coupled to input and output terminals. Re-entrant rods are used for tuning, and the mutual couplings are determined by proper spacing of the individual stubs, and by the use of an appropriate number of rods between the stubs. Fig. 7 is a plot of the reflection coefficient at the input terminal with the output terminated in a resistive load of 51.5 ohms. The peak amplitude of the ripples across channel five was adjusted to be 3%, and the measured reflection coefficient on channel four is of the order of 99% or greater. It can be seen that the bandpass filters used closely meet the ideal requirement of zero reflection on channel five and 100% reflection on channel four.

The Antenna Unit

The final combining of the channel four aural signal with all other signals takes place in the antenna unit in Fig. 8, and shown schematically in Fig. 9. The end terminal of a bridge, Fig. 10, is fed with all signals previously combined in the notch diplexer and combining unit, and these signals therefore appear out-of-phase at the bridge output terminals. The channel four aural signal is brought through a series connected resonant stub assembly, designed to pass only a small band of frequencies centered about the aural carrier frequency, to junction point *A*, and thence to the side input of the bridge. This signal will therefore split into two equal in-phase components at the bridge output terminals. From junction *A*, a separate transmission line leads to an absorbing resistor, and at point *C*, a shunt connected resonant stub, tuned to channel four aural carrier frequency, is connected across the line. Length *AC* is made one quarter wave long in order that the channel four aural signal will see infinite impedance looking towards the absorbing resistor, and will deliver its entire power to the bridge.

As has been previously mentioned the antenna has a significant reflection on channel five. The returning reflected signals arrive at the bridge output terminals in-phase, and therefore appear at the side input of the bridge, where they are directed towards the absorbing resistor. Length *AB* is chosen so that a channel five signal looking from *A* towards *B* sees an infinite impedance, and since the stub connected at *C* has a very high reactance at channel five frequencies, substantially all the channel five energy reflected from the antenna is absorbed, and therefore radiation of a ghost signal is effectively prevented.

Conclusions

It has been established through the successful application of the quadruplexer described herein, that simultaneous operation of two television transmitting stations of medium power into a common antenna and feed system is practical. In the case of CFCM-TV and CKMI-TV substantial savings resulted from the use of the quadruplexer, rather than separate antenna systems.

It has been further shown that, in the case of quadrature fed antenna systems employing dual transmission line feeders, the antenna impedance match need meet usually accepted standards on only one of the two channels. Reflection absorption within the quadruplexer effectively prevents radiation of a ghost signal on the second channel.

While extensive use in television broadcasting of filter systems employing principles similar to those of the quadruplexer described is not envisaged at the present time, it is likely nevertheless that such systems may prove distinctively advantageous and worthy of careful consideration in certain specific cases.

A Transistorized Compandor

J. C. PERKINS, JR.
ASSOCIATE MEMBER AIEE

D. A. PERREAULT
NONMEMBER AIEE

A. F. PERKINS
NONMEMBER AIEE

THE APPLICATION of multiple-channel communication circuits as provided by carrier and microwave radio systems has been steadily increasing in recent years in the attempt to satisfy the widespread consumer demands, public, private, and military included, for more and better telecommunication facilities. The use of these systems has solved many problems of expansion; however, in turn, the systems have introduced their own problems, in particular the increased susceptibility to interference, crosstalk, and noise as compared to physical voice frequency circuits.

In addition, many existing physical circuits will no longer meet today's toll transmission standards for crosstalk and noise due to processes of a slow, deteriorating nature. The addition of individual compandors to such circuits, either physical or carrier-derived, can in many cases prove economically feasible as a method of improving these facilities to a suitable quality.

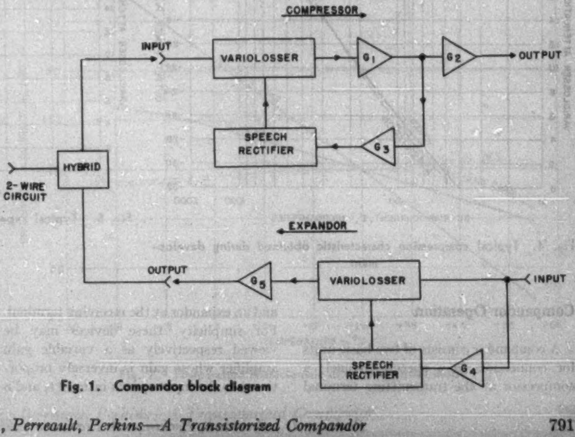


Fig. 1. Compandor block diagram

The principle of companding (compressing and subsequent expanding of the dynamic speech range) has been in use for many years by the telephone industry as a method of improving the transmission quality of communication circuits;[1,2] however, only in recent years has attention been given to the design of versatile, self-contained compandors which can be adapted to a wide variety of application.[3]

This paper describes the design of the Stromberg-Carlson type *545* compandor for general use in voice transmission communication circuits.

Paper 58-906, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 19, 1958; made available for printing May 5, 1958.

J. C. PERKINS, JR., is with the Stromberg-Carlson Company, Rochester, N. Y.; and D. A. PERREAULT and A. F. PERKINS are with O'Hara Associates, Syracuse, N. Y.

The authors wish to acknowledge the initial contributions to this project given by J. L. Wheeler and T. F. Bryzinski of the Stromberg-Carlson Research Department, and the assistance of B. R. Stachiewicz, formerly of the Stromberg-Carlson Company, who established the performance objectives and co-ordinated the development program.

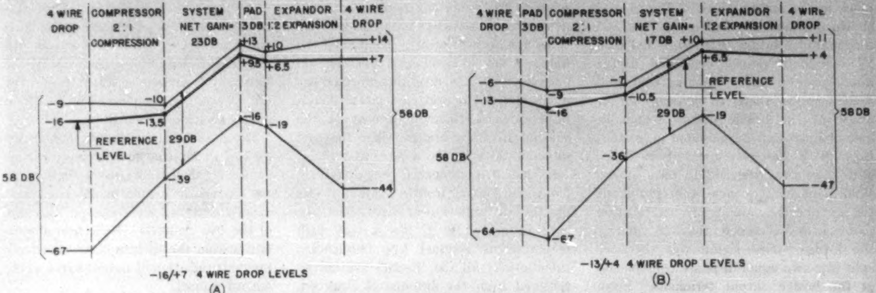

4 WIRE DROP
COMPRESSOR 2:1 COMPRESSION
SYSTEM NET GAIN= 23DB
PAD 3DB
EXPANDOR 1:2 EXPANSION
4 WIRE DROP
+13
+10
+14
+9.5
+6.5
+7
-9
-10
-16
-13.5
-16
-19
58 DB
REFERENCE LEVEL
29DB
-39
58 DB
-44
-67
-16/+7 4 WIRE DROP LEVELS
(A)
4 WIRE DROP
PAD 3DB
COMPRESSOR 2:1 COMPRESSION
SYSTEM NET GAIN= 17DB
EXPANDOR 1:2 EXPANSION
4 WIRE DROP
+10
+11
+6.5
+4
-6
-7
-13
-9
-10.5
-16
REFERENCE LEVEL
29DB
-19
58DB
58 DB
-36
-47
-64
-67
-13/+4 4 WIRE DROP LEVELS
(B)

Fig. 2. Level diagram showing compression and expansion ranges

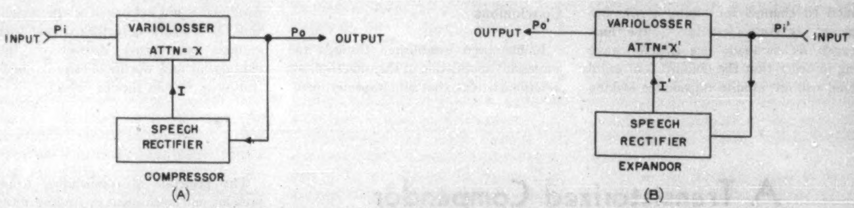

INPUT
Pi
VARIOLOSSER
ATTN= X
Po
OUTPUT
I
SPEECH RECTIFIER
COMPRESSOR
(A)
OUTPUT
Po'
VARIOLOSSER
ATTN= X'
Pi'
INPUT
I'
SPEECH RECTIFIER
EXPANDOR
(B)

Fig. 3. Basic compressor and expandor relationships

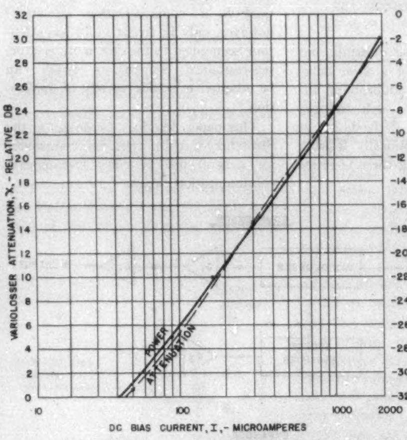


Fig. 4. Typical compression characteristic obtained during development

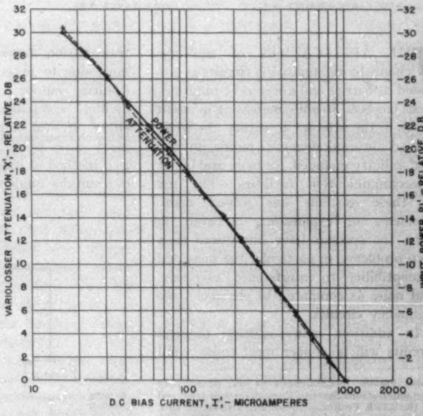


Fig. 5. Typical expansion characteristic obtained during development

Compandor Operation

A compandor consists of two basic units for connection to a speech channel: a compressor at the transmitting terminal and an expandor at the receiving terminal. For simplicity, these devices may be viewed respectively as a variable gain amplifier whose gain is inversely proportional to the input speech intensity, and a variable attenuator whose loss is directly proportional to its input. Thus high-level speech signals receive little gain or loss in passing through both compressor and expandor, while low-level signals are raised in the compressor and subsequently depressed in the expandor to their original

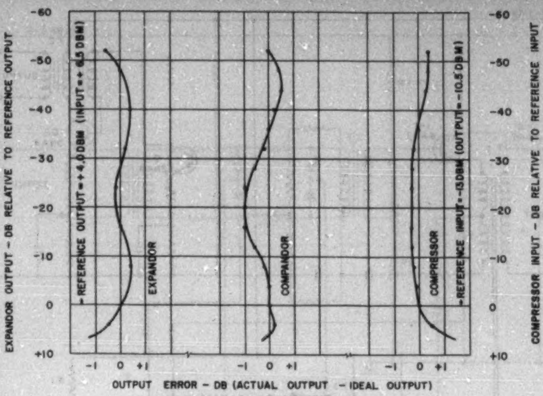


Fig. 6. Tracking characteristic

intensity. This dynamic action of the compandor, when designed to occur at a syllabic rate, will provide a listener with an apparent signal-to-noise improvement in the order of 22 db (decibels) for conventional compandors having compression ratios of 2 to 1. This improvement can vary considerably with speech volume and noise level.[3]

The *545* compandor is of the conventional type: its originality lies in the complete use of transistors as active elements, compact modular construction, and printed circuit techniques.

Design Objectives

In the development of a new product for telephone applications, features which must be considered of primary importance include reliability, power sources, size, installation, and maintenance.

To an industry which is accustomed to measuring performance of a device in terms of 20-year life, the introduction of the transistor will undoubtedly be the main factor in improving the reliability of electronic telephone equipment. The hot cathode vacuum tube, in spite of its extreme versatility and usefulness, is inherently a device of limited life. From all presently available information on transistor life, this limitation does not appear to exist in properly designed semiconductor devices. For this reason it was decided to design the compandor using transistors throughout.

Transistors are also well suited for operation from the standard 48-volt telephone central-office battery. Vacuum tubes usually require a higher voltage supply, which necessitates the additional design of a-c to d-c converters or the installation of a secondary battery supply. The low total power consumption of transistors is still another point in favor of their usage.

Within the past decade, telephone operating companies have been forced to increase their inside plant facilities to keep up with the service demands of expanding population. Thus, physical size is becoming increasingly important in the design of new products. Transistors, in conjunction with miniature components and printed circuit boards, are readily adaptable in providing a compact design.

Maintenance costs consume a large portion of an operating company's budget. Although increasing the reliability of equipment will tend to reduce these costs, it is still necessary to consider the servicing aspects of new designs. For this reason, it was felt that plug-in construction of the compandors was essential to keep the out-of-service time of compandored channels to a minimum.

Electrical Arrangement

In Fig. 1, the electrical arrangement of the compandor is shown in block form. The basic unit is designed for 4-wire operation, although resistance hybrids are included as strapping options where 2-wire operation is required. Input signals to the compressor pass through a variable gain device commonly referred to as the variolosser, are amplified and passed to the line. The variolosser is d-c controlled by a rectified portion of the output speech in a backward-acting circuit, its loss increasing as the intensity of the input signal increases, thus reducing the dynamic range of the output.

In the expandor, compressed signals received from the far terminal are impressed on a variolosser controlled in a forward-acting circuit. As the input increases, the loss is reduced, thus increasing the dynamic range of the output.

The compressor and expandor actions for a single test tone of varying intensity are graphically illustrated in Fig. 2. Although compandors can be applied to a variety of operating levels, the two cases shown are for level points of $-16/+7$ and $-13/+4$ db referred to 1 milliwatt which are most commonly encountered in tele-

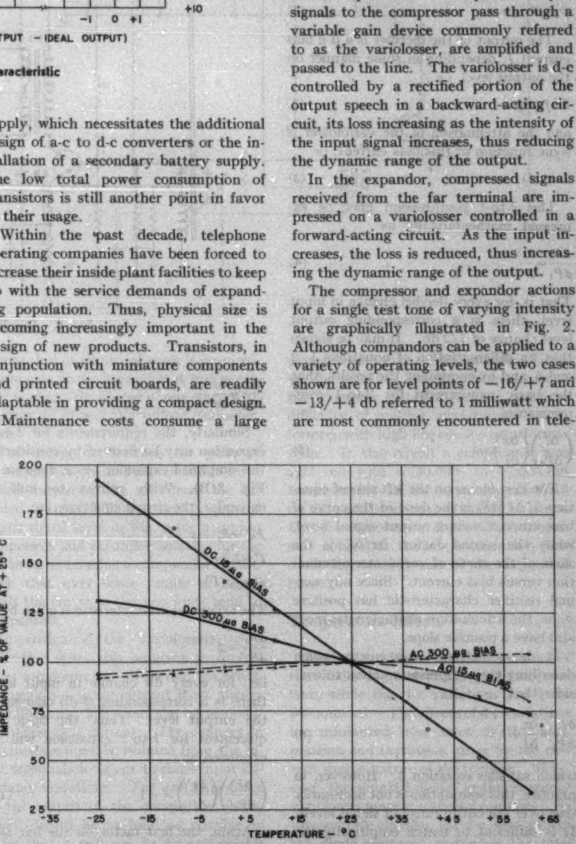


Fig. 7. Temperature characteristic of the variolosser diode

phone offices. It will be noted that the input levels to both compressor and expandor are made identical for both cases by the use of 3-db strap pads. Internal gain controls are then used to provide proper drop levels and to accommodate the required net system gain.

Variolosser Design

To develop the requirements for 2-to-1 compression the more simple compressor block diagram in Fig. 3(A) may be considered, since the amplifier gains are all constant and will not enter into the final result. Expressing the relationship between input power P_i, output power P_0, and attenuation x, in db

$$P_0 = P_i - x \tag{1}$$

The output of the rectifier is a d-c bias current I, depending in some manner on the output level, i.e.,

$$I = f_1(P_0) \tag{2}$$

and the attenuation of the variolosser is some function of this bias current, or

$$x = f_2(I) \tag{3}$$

The 2-to-1 compression may be expressed mathematically as

$$\frac{dP_0}{dP_i} = \frac{1}{2} \tag{4}$$

That is, for every decibel change in input level, there is a corresponding output level change of 1/2 db. It may be shown that by differentiation of equations 1, 2 and 3, and by combining the results, that the basic requirement for 2-to-1 compression is

$$\left(\frac{\partial I}{\partial P_0}\right)\left(\frac{\partial x}{\partial i}\right) = 1 \tag{5}$$

The first factor on the left side of equation 5, $\partial I/\partial P_0$ is the slope of the curve of bias current versus output signal level, while the second factor, $\partial x/\partial I$, is the slope of the curve of variolosser attenuation versus bias current. Since any normal rectifier characteristic has positive slope, the attenuation characteristic must also have a positive slope.

It is sometimes convenient when describing the compression action to consider that

$$\frac{\partial I}{\partial P_0} = \frac{\partial x}{\partial I} = 1 \tag{6}$$

which satisfies equation 5. However, in practice, this assumption is not necessarily true over the entire range of signal levels. It is sufficient to match empirically the slopes of the functions as plotted in Fig. 4, which satisfies equation 5 in the form

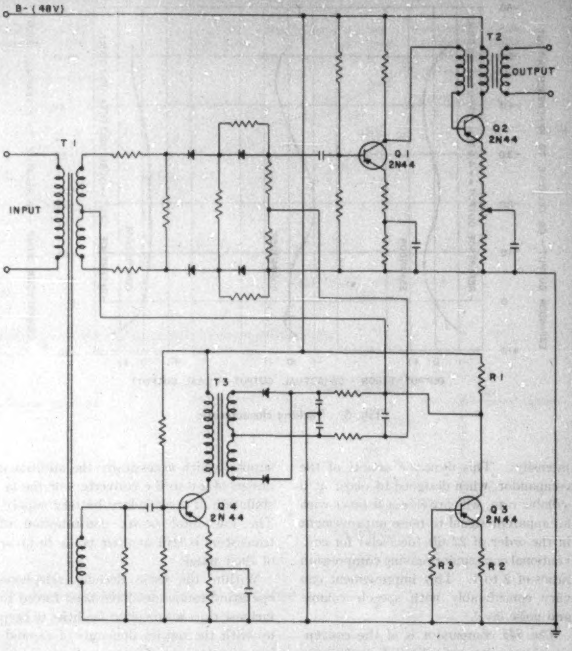


Fig. 8. Simplified expandor schematic

$$\frac{\partial I}{\partial P_0} = \frac{\partial I}{\partial x} \tag{7}$$

Similarly, the requirements for 1-to-2 expansion may be derived by considering the simplified expandor block diagram of Fig. 3(B). With primes to indicate expandor, the circuit equations are

$$P_0' = P_i' - x' \tag{8}$$

$$I' = f_1'(P_i') \tag{9}$$

$$x' = f_2'(I') \tag{10}$$

The expansion characteristic is given by

$$\frac{dP_0'}{dP_i'} = 2 \tag{11}$$

i.e., for every db change in input level there is a corresponding 2-db change in the output level. Thus, the basic requirement for 1-to-2 expansion will be expressed as

$$\left(\frac{\partial I'}{\partial P_i'}\right)\left(\frac{\partial x'}{\partial I'}\right) = -1 \tag{12}$$

Again, the first factor on the left side of equation 12, $\partial I'/\partial P_i'$, is the slope of the curve of bias current versus input signal level while the second factor, $\partial x'/\partial I'$, is the slope of the curve of variolosser attenuation versus rectifier current. This quantity is made negative by means of the variolosser circuit configuration, i.e., the attenuation decreases as the current increases.

The design approach is similar to that for the compressor. The customary plot of the functions is as shown in Fig. 5. By empirically matching the slopes of these curves equation 12 may be satisfied in the form

$$\frac{\partial x'}{\partial I'} = -\frac{\partial P_i'}{\partial I'} \tag{13}$$

In both the compressor and expandor, the desired rectifier characteristic is approached by designing the rectifier as a current source so that the effect of load changes caused by variations in the d-c resistance of the variolosser with current will be minimized. This approach is limited because of the wide range of load variation which the variolosser presents and the relatively small amount of power available at the input to the rectifier.

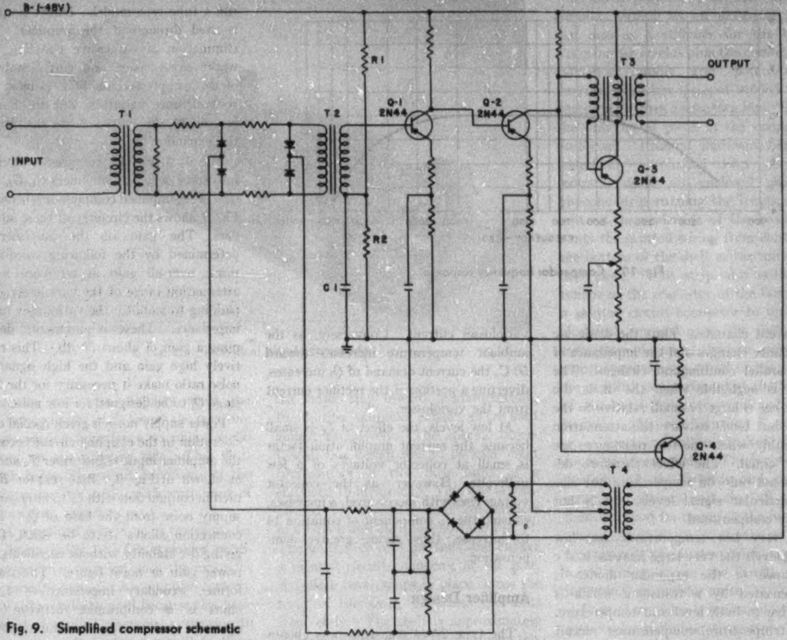


Fig. 9. Simplified compressor schematic

The variolosser is a balanced attenuator whose attenuation depends on the bias current fed to it from the rectifier by means of a simplex-type connection. The positive sloping attenuation function required in the compressor is obtained by using series resistance followed by shunt diodes. The negative slope required in the expandor is obtained by using series diodes followed by shunt resistors.

The approximate diode characteristic appropriate for this type of circuit is generally well known. The exact characteristic was determined in conjunction with diode manufacturers experienced in supplying this type of diode, so as to take advantage of their maximum yield. The diodes are selected on the basis of their dynamic impedance when biased in the forward direction.

With this type of diode, a 2-stage variolosser is required in both the compressor and the expandor in order to obtain the wide range of attenuation required. Exact matching of the characteristics shown in Figs. 4 and 5 is accomplished by adjusting the variolosser and rectifier resistors. Since these circuits are not entirely independent of each other, a great deal of empirical work is required. In the case of the expandor, it was necessary to shunt the diodes of one stage with resistors to obtain the proper characteristics.

The difference between the pairs of curves in Figs. 4 and 5 represents an error in the output of the unit from the expected output. This error is plotted versus signal level in Fig. 6 for a typical compressor and expandor and also for the two in series forming a compandor circuit. Note that over some ranges of signal levels the errors in the two units tend to compensate.

In designing the variolossers, consideration must also be given to several other requirements. To avoid distortion the peak a-c signal current must always be sufficiently less than the corresponding diode bias current. The changes in attenuation must be isolated from the input terminals so as not to cause input impedance variations. The load impedance at the output of the compressor variolosser should be as high as possible in order that it will not be an appreciable shunt on the diodes at the low attenuation (high diode impedance) end of the compression range. The limitations here are the relatively low input impedance of the transistor amplifier and the practical difficulty of making miniaturized transformers with high-impedance transformations. In this circuit a compromise was achieved with a 30,000- to 1,200-ohm transformer. In the expandor the load on the variolosser output should be low so that the diodes can still provide the necessary control at the low attenuation (low diode impedance) end of the expansion range.

Temperature Effects

As shown in Fig. 7, the a-c impedance of the diodes is quite stable with temperature, while the d-c variations are quite pronounced. The effects of d-c variations are minimized over most of the compression and expansion range by the current source nature of the rectifier. However, the shunted diodes in the expandor variolosser present a special problem. As temperature changes, the d-c resistance of the diode changes; and while the total current through the parallel combination remains relatively constant, the division

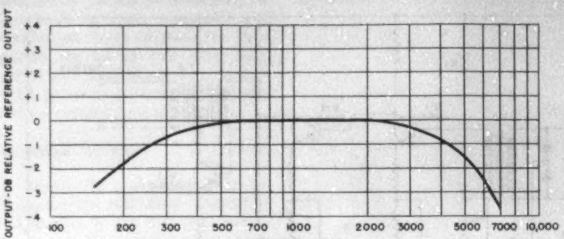


Fig. 10. Compandor frequency response

of current changes. Thus the diode a-c impedance changes and the impedance of the parallel combination changes. The effect is negligible when the diode d-c resistance is large or small relative to the shunt but tends to vary the attenuation noticeably when the two resistances are about equal. The effect therefore depends not only on temperature but also on particular signal levels and is not readily compensated.

At very low temperatures and low signal levels the very large increase in d-c resistance of the expandor diodes is compensated by a transistor which is sensitive to both level and temperature. The temperature compensating circuit comprises Q_3, R_1, R_2, and R_3 of Fig. 8. The circuit controls the direction and magnitude of current flow between the collector of Q_3 and the variolosser as a function of temperature, primarily, and with level as a secondary effect. The current flowing in R_1 is determined by $B-$ since the collector is close to ground. The current required by Q_3 is determined by the following well-known equations:

$$I_C=\alpha I_E+Ico \qquad (14)$$

$$I_E=E_B/R_2 \qquad (15)$$

$$E_B=Ico(R_3) \qquad (16)$$

Since *Ico* is a function of temperature, I_c is temperature dependent, and its shape is determined by the values of α, *Ico*, R_2, and R_3.

The values of R_1, R_2, and R_3 are so chosen that the collector current required at 25 C (degrees centigrade) is approximately equal to the current supplied from $B-$ through R_1. At this temperature there is no current flow between Q_3 and the variolosser. The entire variolosser current is supplied by the rectifier. As the ambient temperature approaches zero C, the current demand of Q_3 decreases, thus allowing a portion of the current through R_1 to supplement the variolosser current. Conversely, as the ambient temperature increases toward 50 C, the current demand of Q_3 increases, diverting a portion of the rectifier current from the variolosser.

At low levels, the effect of I_E is small because the current amplification factor is small at collector voltages of a few millivolts. However, as the collector voltage rises with speech level, α increases, causing the I_E component of equation 14 to increase, thus giving greater compensation.

Amplifier Design

The type *2N44* transistor was chosen as a medium beta relatively high peak inverse voltage unit, allowing economical use of the $B-$ supply. Solid tantalum capacitors are used in the amplifiers because they are inherently stable with temperature, they are considered to have long life, and they are small in size.

The power drain of the compandor has been kept to a minimum consistent with stability and reliability. The unit draws 30 milliamperes at 48 volts or about 1.5 watts of power. This is compared to about a 4.5-watt filament power alone to run a tube-type model. Direct coupling is used throughout the amplifier. The elimination of capacitive coupling networks saves space and cost. Stability of the compressor amplifier is most important since variations due to changes in gain of the amplifier are doubled in the expandor.

Fig. 1 shows the compressor divided into three separate amplifiers G_1, G_2, and G_3. The simplified compressor schematic, Fig. 9, shows the circuitry of these amplifiers. The gain of the amplifier is determined by the following considerations: over-all gain at reference level, attenuation range of the variolosser, and padding to stabilize the variolosser input impedance. These requirements determine a gain of about 50 db. This relatively high gain and the high signal-to-noise ratio make it necessary for the first stage Q_1 to be designed for low noise.

Power supply noise is given special consideration in the coupling circuit between the amplifier input transformer T_2 and Q_1 as shown in Fig. 9. Bias resistor R_1 is used in conjunction with C_1 to filter power supply noise from the base of Q_1. This connection allows R_2 to be small, thus giving d-c stability without sacrificing a-c power gain or noise figure. The transformer secondary impedance of 1,200 ohms is a compromise between the optimum values for maximum power transfer and low noise figures.

Isolation between the rectifier and the signal output is provided by the rectifier amplifier Q_4, and the output amplifier Q_5. The rectifier amplifier is required because the *2N44* transistor is not capable of supplying sufficient power for both the line and the rectifier over the required temperature range.

The change in output impedance of the amplifier with variations in transistor parameters presents an additional problem. The unbypassed emitter resistor is

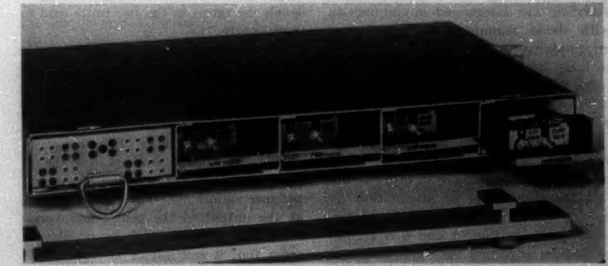


Fig. 11. Compandor shelf mounting four compandors and junction box

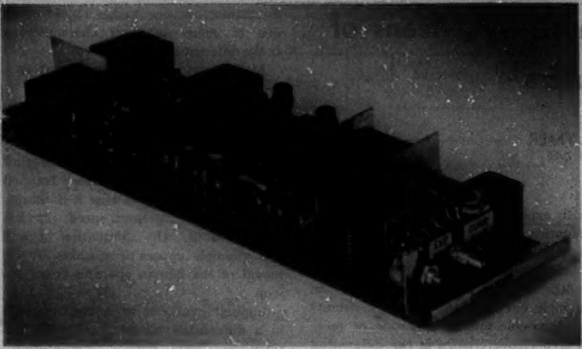


Fig. 12. Individual plug-in compandor

high to give gain stability while the tertiary winding of T_3 lowers the resulting high output impedance to about 4,000 ohms on the primary side. This feedback holds the output impedance of the compressor to ±5% with variations of transistor parameters.

A simplified schematic of the expandor is shown in Fig. 8. The expandor amplifier design is simpler since the expandor is forward-acting. An amplifier preceding the rectifier is used because this basic design is applied to special systems with lower expandor input levels than those shown in Fig. 2.

For economy of manufacturing, the same rectifier transformer and output transformer are used in both compresser and expandor. A full-wave bridge is used in the compressor where the maximum d-c resistance of the load on the rectifier is greatest because no shunts are used on the variolosser diodes. The extra diodes are relatively inexpensive and are in fact of the same type used in the variolosser, without special selection.

Fig. 10 shows the compandor frequency response.

Mechanical Design

The compandor is a modular, printed circuit design. The use of transistors was the major step toward miniaturization. In addition, tantalum capacitors are used almost exclusively. The specific objective in this design was to package eight compandors in two rack units of a standard 19-inch rack. This was accomplished by mounting four compandors in a 1-rack unit shelf, as shown in Fig. 11. The units slide into the shelf and engage a printed circuit connector at the rear. A 1-piece cover snaps in place across the front of the shelf to enclose the units completely. The shelf is approximately 10 inches deep and extends approximately 6 inches beyond the front face of the mounting rack.

To increase the volumetric efficiency and still maintain simplicity and accessibility, modular assembly of the smaller components was adopted. The modules and larger components such as transformers are mounted on the main printed circuit board. The "decks" of the modules are also printed circuits. In general, the modules represent logical circuit groupings; variolosser, amplifier, etc. Since the compandor is made up of two completely separate circuits, each with wide ranges of signal levels, careful attention was given to the physical layout. The compressor is located along the right side of the unit, as shown in Fig. 12, and the expandor along the left side.

A metal partition separates the two circuits, serving as an electrical shield and also as a stiffener for the board. Transformer shields plus the combination extractor handle and potentiometer mounting bracket are spot-welded to the partition to form a subassembly. This is the only metal piece in the compandor unit itself. Framing hardware has been completely eliminated since the unit mounts inside an enclosed shelf. A phenolic strip running the length of the unit on the underside of the board prevents the printed wiring from contacting the bottom of the shelf as the unit is slid into place. This strip also guides the tongue at the rear edge of the board into a printed circuit connector to make the electrical connections to the shelf. The connectors of each of the four compandor compartments were wired to the connector of a fifth compartment at the left side of the shelf by a conventional wiring harness. This compartment is occupied by a "junction box" to which all external connections are made. The junction box contains all the hybrids and pads for strapping options and also all the tip jacks required for metering the input and output levels of the compandors. This arrangement eliminates much back-and-forth wiring across the length of the individual compandor units.

Ease of servicing is increased by use of sockets for mounting transistors and by the special manner in which the modules are connected to the printed board. The bus wire studs of each module are dip-soldered to terminals swaged to the circuit side of the main board, rather than soldered to the printed wiring itself, in order that replacement may be accomplished without damage to the printed board.

References

1. The Compandor—An Aid Against Static in Radio Telephony, **R. C. Mathes, S. B. Wright.** *Bell System Technical Journal*, New York, N. Y., vol. 13, 1934, pp. 315–32.

2. Application of Compandors to Telephone Circuits, **C. W. Carter, Jr., A. C. Dickieson, D. Mitchell.** *AIEE Transactions*, vol. 65, 1946, pp. 1079–86.

3. A Miniature Compandor for General Use in Wire and Radio Communication Systems, **F. S. Boxall, R. S. Caruthers.** *Ibid.*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 804–11.

Direct System Synthesis by Means of Computers

A. BEN CLYMER
NONMEMBER AIEE

BEFORE a system can be simulated on an operational analog computer, a passive-element analog computer, a digital differential analyzer, or a digital computer, it is necessary to derive or synthesize a set of differential equations which yield the response of the system to an arbitrary excitation. Every constant and function in the differential equations must be known.

For some systems the state of the art permits one to derive theoretically a formula or value for each function and constant. It is then straightforward to simulate such a system on a computer.

Most systems, however, have at least one constant or function which cannot as yet be obtained from theory. For these systems it is necessary to have recourse to experimental determination. If experimental techniques can be devised for the measurement of the "characteristics" (i.e., coefficients, exponents, functions, etc.) which cannot be obtained from theory, then analog simulation of even these systems is straightforward.

The systems which can be treated by the methods considered in this paper are those which do not readily reveal all of their characteristics when investigated by means of both theory and experiment. They are legion.

This paper presents two general methods for using a computer in conjunction with experimental results for the determination of characteristics in the differential equations representing a system. One method, called herein "implicit synthesis," consists of the use of an implicit circuit (i.e., "forcing" circuit) to determine and display one or more desired characteristics in the differential equations for a system, given the rest of the terms in the differential equations, as well as the dynamic response of the system to some arbitrary excitations which exercise the dependent variables over their full ranges. The implicit circuit contains a mechanization of a differential equation which is complete except for the unknown characteristics. The missing characteristics are synthesized by the implicit circuit by continuously comparing the values generated for dependent variables with the experimental time histories of the dependent variables, and greatly amplifying the differences. A more extended explanation of the principle will be offered later.

Implicit synthesis is thus a distinct mode of application of computers. It differs from the usual mode of application, in which one knows the differential equation, including all characteristics, and one knows the excitation, while one wishes to determine the solution to the equation, which is the dependent variable as a function of the independent variable. By contrast, in implicit synthesis one knows the solution and the excitation, but one wishes to determine one or more unknown characteristics in the differential equations.

Much more common at present as a technique than implicit synthesis is manual trial-and-error setting of the potentiometers in an explicit analog circuit in order to match by eye or by minimum error criterion a given time history of a dependent variable. Implicit synthesis is superior because it is characterized by: 1. greater speed, 2. less error, and 3. the elimination of several possible causes of mistakes by removal of the human operator from the loop.

Implicit synthesis of a characteristic in a differential equation is to be distinguished from "implicit computation" of a quantity from an algebraic relationship.[1] A familiar example of an implicit computation is division by means of a multiplier and a high-gain amplifier. In this example, one wishes to find y and $xy=z$, given x and z. It can be done by means of a circuit of the type diagrammed abstractly in Fig. 1.

Extensive inquiry on the part of the author has failed to reveal any significant use of implicit synthesis on general purpose computers. It is apparently a virgin field, so far as most users of computers are concerned. Yet implicit synthesis has an interesting history which dates at the latest from World War I, when the Ford Instrument Company employed this principle in a mechanical analog device for determining the speed and course of a target ship for naval fire control purposes, where the range and bearing to the target were known continuously and one's own ship's speed and course were known. More elaborate applications have since been developed in the field of fire control engineering, and in other branches of the special-purpose computer field as well.

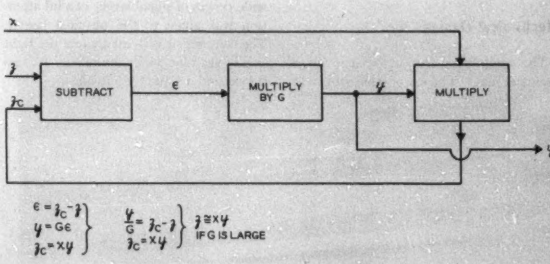


Fig. 1. Division by implicit computation

Paper 58-1003, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted May 8, 1958; made available for printing June 11, 1958.

A. BEN CLYMER is with the Columbus Division of North American Aviation, Inc., Columbus, Ohio.

For the analog computer findings reported herein and for valuable suggestions concerning the design of circuits for implicit synthesis, the author is indebted to Arnold M. Bartz of Dow Chemical Company, Midland, Mich., formerly with North American Aviation, Inc.; John J. Nichol, James C. Italiano, Raymond M. Kisner, and Charles R. Walli, of the Columbus Division of North American Aviation, Inc., Columbus, Ohio. Burt Went, also of the Columbus Division, conducted the digital computer studies. The author acknowledges also the many ideas derived from conversations with Vernon Larrowe of the Willow Run Research Laboratory of the University of Michigan, Ypsilanti, Mich.; David Sparks of Tammen and Denison, Chicago, Ill.; Howard Hamer and Jerome Kennedy of Electronic Associates, Inc., Long Branch, N. J., and Chicago, Ill., respectively; and others.

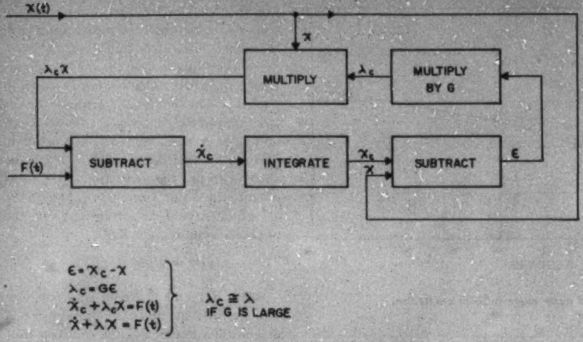


Fig. 2. Implicit synthesis of first-order system

Implicit Synthesis

Implicit Synthesis of First-Order System

If one has a differential equation, say, $\dot{x}+\lambda x=F(t)$, in which one knows $x(t)$ and $F(t)$ but wishes to determine λ, the implicit computation is called herein "implicit synthesis of the characteristic λ." An abstract diagram of a circuit for this purpose is given in Fig. 2. This is a simple case of implicit synthesis; accordingly, the circuit will be explained.

In Fig. 2, $x(t)$ and $F(t)$ are given functions of time, and λ_c is the answer computed by the circuit. Consider the consequences if λ_c were not very nearly correct: $\lambda_c x$ would be wrong, and hence so would $\dot{x}_c$, which would make x_c start to vary away from its nearly correct value, thus making ϵ vary, and this in turn would change λ_c. If the algebraic signs in the circuit are properly designed, λ_c will be driven by the forcing circuit to very nearly its correct value, in principle. If $F(t)$ comes to a constant steady-state value, λ_c will become constant at very nearly its correct value and x_c will very nearly equal x. Even if $F(t)$ varies, the circuit will strive to maintain λ_c at very nearly its correct value at all times. The larger the value of the gain G, the smaller the error in λ_c, provided that the circuit remains stable.

Of course, no one would trouble to use a computer to determine λ in this case, since it would be so simple to calculate or to measure on the actual system a static x under a steady F and calculate λ as F/x. Nevertheless, the circuit will be discussed here in some detail, since it embodies many of the principles which are involved in determining system characteristics in complicated sets of nonlinear differential equations which do not lend themselves readily to experimental determination on the actual system.

One may analyze the behavior of the circuit shown in Fig. 2 by manipulation of the equations which represent the circuit and the physical system:

$$\epsilon=x_c-x \tag{1}$$

$$\lambda_c=G\epsilon \tag{2}$$

$$\dot{x}_c+\lambda_c x=F(t) \tag{3}$$

$$\dot{x}+\lambda x=F(t) \tag{4}$$

where

ϵ = error in x_c
x_c = computed value of x at any instant
x = actual response of the first-order system
G = large number (gain)
λ_c = computed value of λ at any instant
λ = unknown characteristic of the first-order system
$F(t)$ = excitation of the first-order system

First, ϵ can be eliminated between equations 1 and 2:

$$G^{-1}\lambda_c=x_c-x$$

$$\text{or}\quad x_c=x+G^{-1}\lambda_c \tag{5}$$

Next, x_c can be eliminated between equations 3 and 5:

$$\dot{x}+G^{-1}\dot{\lambda}_c+\lambda_c x=F(t) \tag{6}$$

Finally, $F(t)$ can be eliminated between equations 4 and 6:

$$G^{-1}\dot{\lambda}_c+x\lambda_c=\lambda x \tag{7}$$

If x is constant, the circuit may be seen to have a damped exponential transient whose time constant is $(Gx)^{-1}$, i.e., the higher the gain or the larger the response x, the more rapidly the circuit attains the desired answer. On the other hand, the lower the gain or the smaller the value of x, the more sluggishly λ_c approaches λ.

It is to be noted that the circuit goes unstable when x goes negative. To prevent this, it is necessary to include in the circuit a device which maintains at all times a set of algebraic signs which keeps the system stable, regardless of the sign of x. This can be accomplished by means of an absolute magnitude circuit inserted either in the λ_c line or in the x line to the multiplier.

The first alternative is used in the analog circuit shown in Fig. 3, which is one possible way to render the abstract circuit in Fig. 2 with operational analog components. Two slight refinements will improve the performance of this circuit. One is to put a capacitor (about 1,000 micromicrofarads) around the high-gain amplifier as a filter and also to reduce the steady-state error. Another is to bias out any unbalance voltage

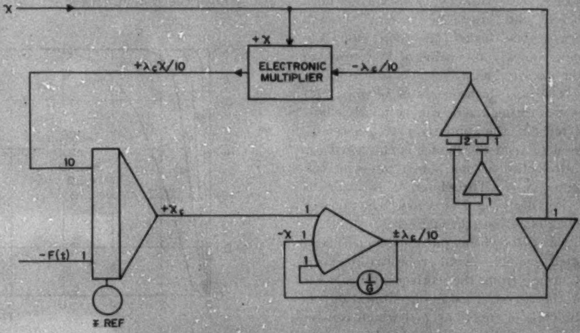


Fig. 3. Analog circuit for implicit synthesis

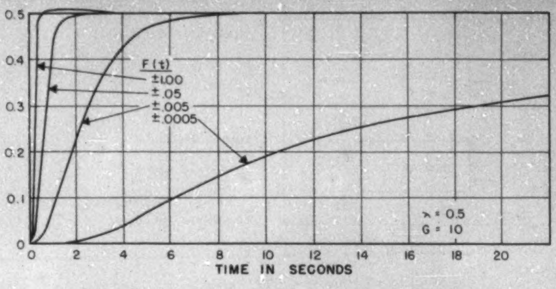


Fig. 4. Dependence of circuit speed upon magnitude of excitation

arising from the diodes by means of a potentiometer fed by reference voltage, in order to insure that the voltage $-\lambda_c/10$ never changes sign, especially for small values of input.

The implicit synthesis circuit shown in Fig. 3, when set up on an Electronic Associates analog computer, did an excellent job of determining λ, provided that x stayed away from a narrow band of values around zero. For example, Fig. 4 shows the behavior of λ_c when x is constant. It is evident that λ_c always approaches the correct value λ with a speed which is greater for larger values of $F(t)$. However, there is a steady-state error which varies inversely with $F(t)$. The error is eliminated if a filter capacitor is used around the high-gain amplifier, because the capacitor introduces an integration. Fig. 5 shows that higher loop gain in the implicit synthesis loop does increase the speed with which the answer is approached, as derived by the foregoing analysis. If $F(t)$ passes through zero, the circuit momentarily generates a wrong answer, as shown in Fig. 6 (A). However, λ_c is correct both before and after the passage through zero. The higher the loop gain, the more quickly the value of λ_c returns to the correct value λ.

The pip which occurs in λ_c when $F(t)$ passes through zero can be nearly eliminated if the potentiometer normally used to trim the diodes is employed for introduction of one's best guess of the value of λ. The closer λ_c is made to λ initially, the less wrong the answer when $F(t)$ passes through zero.

An $F(t)$ which passes through zero often enough could conceivably keep λ_c away from the correct value λ all of the time. This is one limitation on the class of functions $F(t)$ which can be used in conjunction with this circuit for implicit synthesis of the characteristic λ of a first-order system. An allowable sinusoidal $F(t)$ is shown in Fig. 6 (B).

If the true λ of the system is suspected of being a function of x rather than constant, this can be tested by plotting λ_c against x instead of against time. If the resulting curve is a horizontal straight line, except for noise and erratic displacements due to shortcomings of the signals and of the implicit synthesis circuit, then λ is a constant. On the other hand, if $\lambda_c(x)$ retraces itself over some curve other than a horizontal straight line, as x varies, then λ has been shown to be the function of x which the computer has plotted as $\lambda_c(x)$. Thus, implicit synthesis can yield nonlinear characteristics as well as constant coefficients.

The accuracy of implicit synthesis of λ for a first-order system, with the use of the analog circuit just discussed with a gain of at least 100, is usually excellent, the error being less than 1%. Additional error could be introduced by the transducers or magnetic tapes which convey $x(t)$ and $F(t)$ to the computer from the test setup of the system. If the system is not in fact exactly first-order, then of course a first-order representation will be still less accurate.

Another approach with which to determine λ for a first-order system is shown in Fig. 7. The circuit determines λx as the amplified difference between x_c and x, and then employs an open-loop division by x to determine λ. It can be shown from the equations for this circuit that the time constant of the λ_c exponential-decay transient in the implicit circuit is given by

$$\tau = (G + \dot{x}/x)^{-1} \tag{8}$$

Clearly, if $\dot{x}$ and x are of opposite sign at some instant, and if their ratio has a magnitude exceeding G, the time constant can go negative. This means that under some conditions the implicit circuit can go unstable. This defect of the circuit makes it unsuitable for some class of functions $F(t)$, just as was the case with the circuit previously discussed.

A further minor difficulty is experienced if the divider is so designed that it must be fed a positive divisor. This would require an absolute magnitude circuit. Also, since "λ_c" changes sign, if x changes sign, either the dividend or the quotient must pass through an absolute magnitude circuit in order to keep the answer of constant algebraic sign.

If λx is plotted against x, one may see whether this function is nonlinear. If it is, usually one can dispense with the divider and its associated absolute magnitude circuits, since then λx is the quantity which is of direct interest. However, one may plot λx from the previous circuit also, so this is no advantage for either circuit.

Preliminary tests indicated no great differences between the performances of the two circuits under most condi-

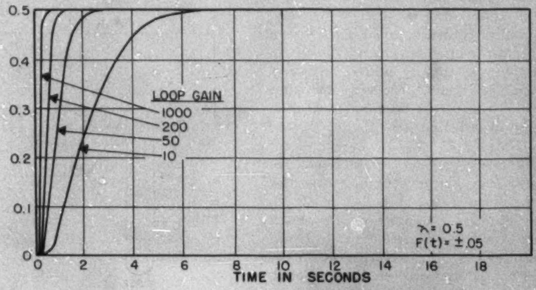


Fig. 5. Dependence of circuit speed upon implicit loop gain

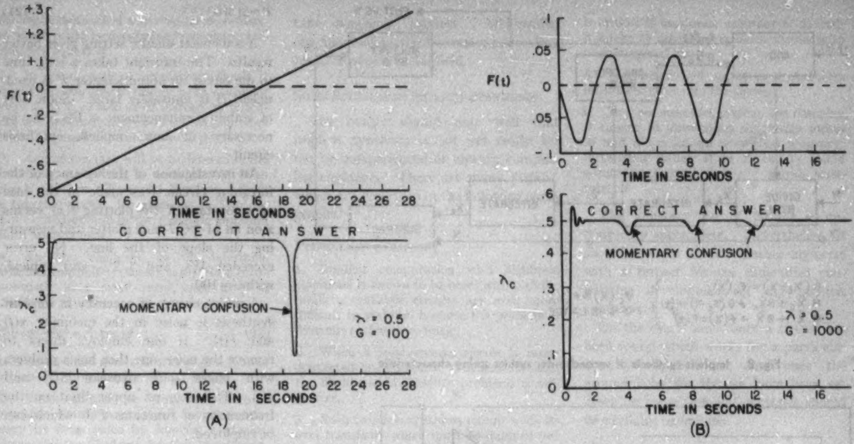


Fig. 6. Two examples of momentary confusion of circuit after crossing zero

tions. The two circuits can be contrasted on the basis of what is in the implicit loop. In one case the implicit loop is chiefly a differentiator, and various operations are then performed open-loop on the result. In the other case the inverses of these operations are performed inside the loop, and the loop yields the answers directly. Any implicit synthesis circuit may be rendered in either of these extreme forms or any form in between (some operations being inside the loop and others open-loop). In some applications it may be desirable to provide automatic loop gain control.

It might be possible to improve the stability of the foregoing circuit by changing the high-gain amplifier to an integrator with large input gains and by feeding the output back to the input through a potentiometer. This modified implicit synthesis circuit would have a second-order instead of a first-order transient. Hence the transient could be damped out quickly by proper setting of the potentiometer, which introduces artificial damping. A corresponding change could be made in the "circuit" if implicit synthesis were conducted on a digital computer or digital differential analyzer.

IMPLICIT SYNTHESIS OF A SECOND-ORDER SYSTEM

Consider next a system of second order of the form

$$M\ddot{x}+B\dot{x}+Kx=F(t) \tag{9}$$

A typical problem for implicit synthesis would be to determine K, knowing M, B, $x(t)$, and $F(t)$.

One approach would be analogous to that used in the foregoing for finding λ for a first-order system, namely, to let the unknown characteristic K be the quantity obtained by amplifying the difference between $x(t)$ and the implicitly synthesized quantity $x_c(t)$. However, it will be more instructive to discuss another approach, which was also applied in the foregoing to a first-order system.

Let the quantity $G(x_c-x)$ represent $Kx=f(x)$ instead of only K. Then this Kx can be plotted against x and the slope can be read off. In fact, if K depends upon x, the function $f(x)$ will be displayed directly as the ordinate of the plot. An abstract circuit for this purpose is given in Fig. 8, and a satisfactory operational analog circuit for implicit synthesis of $f(x)$ is given in Fig. 9. It is to be noted that the circuit in Fig. 9 for a second-order system takes less equipment than the circuit in Fig. 3 for a first-order system.

The circuit in Fig. 9 is easy to analyze in its essential respects. One obtains the following equations:

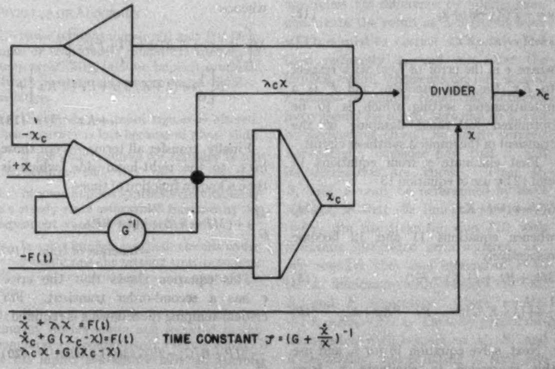


Fig. 7. Alternative analog circuit for implicit synthesis

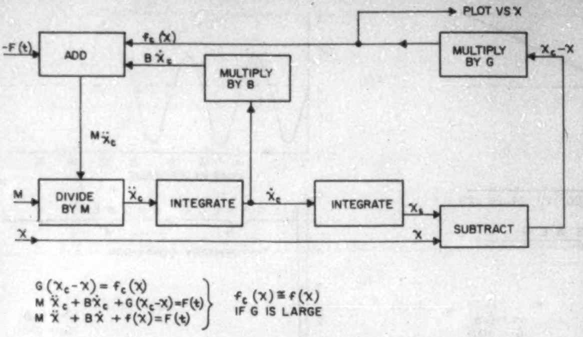


Fig. 8. Implicit synthesis of second-order system spring characteristic

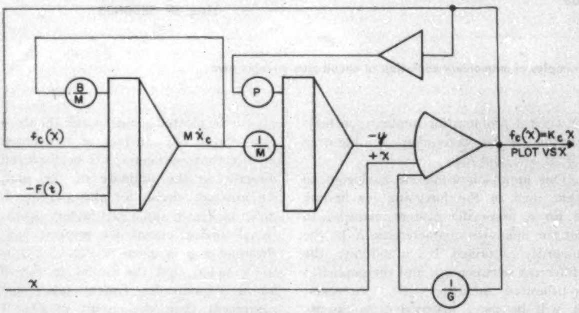


Fig. 9. Analog circuit for finding spring characteristics of second-order system

$$M\ddot{x}+B\dot{x}+Kx=F(t) \tag{10}$$

$$M\ddot{x}_c+B\dot{x}_c+G(y-x)=F(t) \tag{11}$$

$$\dot{y}=-PG(y-x)+\dot{x}_c \tag{12}$$

$$\epsilon=G(y-x)-Kx \tag{13}$$

where ϵ is the error of $f_c(x)$ with respect to Kx, the true function, and P is a potentiometer setting which is to be optimized for rapid damping of the transient of the implicit synthesis circuit.

First eliminate y from equations 11 and 12 by use of equation 13:

$$G(y-x)=\epsilon+Kx \tag{14}$$

whence equations 11 and 12 become respectively

$$M\ddot{x}_c+B\dot{x}_c+\epsilon+Kx=F(t) \tag{15}$$

$$\frac{1}{G}\dot{\epsilon}+(1+K/G)\dot{x}=-P\epsilon-PKx+\dot{x}_c \tag{16}$$

Next, solve equation 16 for $\dot{x}_c$ and use this expression to eliminate x_c from equation 15:

$$\dot{x}_c=\frac{1}{G}\dot{\epsilon}+(1+K/G)\dot{x}+P\epsilon+PKx \tag{17}$$

whence

$$M\left[\frac{1}{G}\ddot{\epsilon}+(1+K/G)\ddot{x}+P\dot{\epsilon}+PK\dot{x}\right]+ B\left[\frac{1}{G}\dot{\epsilon}+(1+K/G)\dot{x}+P\epsilon+PKx\right]+ \epsilon+Kx=F(t) \tag{18}$$

Finally, transfer all terms, except those in ϵ, to the right-hand side, which is then a known function of time:

$$\frac{M}{G}\ddot{\epsilon}+(MP+B/G)\dot{\epsilon}+(1+BP)\epsilon =\text{function of time} \tag{19}$$

This equation shows that the error ϵ has a second-order transient. For critical damping the following is required:

$$\frac{G}{M}(MP+B/G)=2[(G/M)(1+BP)]^{1/2} \tag{20}$$

whence

$$P\approx 2(MG)^{-1/2} \tag{21}$$

A somewhat smaller setting gives better results. The transient takes a long time to die out if no potentiometer P is used, unless B is unusually large. Some sort of stability enhancement is likely to be necessary in any implicit synthesis circuit.

An investigation of the accuracy of the foregoing circuit was made in the case of a constant K by plotting $f(x)$ versus x on an 11- ×17-inch plotter and measuring the slope of the line. No error exceeded 1%, and 1/2% was typical, with $G=100$.

A major enemy of accuracy in implicit synthesis is noise in the incoming $x(t)$ and $F(t)$. If one employs filters to remove the noise, one then has a problem with phase shift between $F(t)$ and $x(t)$. This sets an upper limit on the frequencies of functions $F(t)$ which can be employed.

If the initial value of $f_c(x)$ is not near the correct value, the circuit takes longer to approach the answer. If the value of B is not known accurately, it makes very little difference in the accuracy with which $f(x)$ is determined by the circuit. This is because B has only a minor effect upon the damped natural frequency of a second-order system. In a test of this point it was found that a factor 10 error in B produced no observed change in $f_c(x)$.

NEED OF RESEARCH ON IMPLICIT SYNTHESIS

Only a few circuits have been presented herein for common systems. There are many other systems of practical interest, and each contains several characteristics which could be obtained by implicit synthesis. Moreover, for finding any one characteristic there are many possible variations that one may design into the circuit in the interest of stability, accuracy, and freedom from noise. Several suggestions for needed research are presented in the following.

1. One could aspire to determine more than one characteristic per system equation. For example, one might try to determine B and K for a second-order system, or even M, B, and K. A possible approach would be to feed to the implicit synthesis circuit an additional derivative of the dependent variable for every additional characteristic desired. Another approach to be tried would be to make as many pairs of $F(t)$ and $x(t)$ tapes as there are characteristics to be determined, making sure that the functions $F_i(t)$ are sufficiently "independent" to convey different information, and having an implicit synthesis circuit for each characteristic. A third approach would be to replace an nth order differential equation by

the equivalent set of n first-order equations, in which could be determined one characteristic per equation.

2. In some cases of implicit synthesis it might be desirable to match $\dot{x}_c$ to $\dot{x}$, and/or even $\ddot{x}_c$ to $\ddot{x}$, as well as or instead of matching x_c to x, depending upon which quantities in the system can be measured accurately.

3. A common task will be to determine the functions comprising a term of the form $f(x)\times g(\dot{x})$ in a differential equation. Such a function occurs, for example, in Van der Pol's equation and in the dynamic equation for a landing gear.

4. Another challenge is to determine whether implicit synthesis can yield a characteristic in a small term, such as β in Duffing's equation:

$$\ddot{x}+\alpha x+\beta x^3=A\cos\omega t \tag{22}$$

or a small third derivative term in a system which is approximately second-order.

5. Work should be done to discover whether small nonanalytic nonlinearities may be determined by implicit synthesis. Examples are Coulomb friction, dead band, hysteresis, structural damping, mechanical stops, time delay, and the like.

6. An especially important class of problems today is the development of techniques for writing process equations and determining the characteristics.

7. Studies should be conducted to find out what can be done to supplement implicit synthesis by human trial and error, especially with respect to discovery of the most appropriate form for the differential equations describing a system, or with respect to determining more than one characteristic per equation.

8. Another possibility for research is to discover in what cases implicit synthesis could be accomplished on a digital computer or digital differential analyzer with better accuracy than on an analog computer.

9. Along with research on circuits and techniques, there must be developments in equipment, such as tape recorders and transducers having better linearity and less noise. There might even be a need for a photoelectric curve follower fed by an oscillograph roll. One may think also of the utility of special-purpose portable analog or digital computers for implicit synthesis of system characteristics in the laboratory, simultaneously with the testing of a system.

Implicit synthesis thus is a rich field for research and development studies conducted with computers. In such investigations, all equations and parameters should be nondimensionalized, if maximum economy of effort is desired. No one person or organization can hope to accomplish more than a small fraction of the studies that need to be made. Eventually, however, one may expect that there will be available a large catalog of trouble-free circuits for implicit synthesis of the characteristics of all systems having practical interest or important academic interest. Meanwhile, one may look forward to many interesting research papers in this field.

SOME PITFALLS IN IMPLICIT SYNTHESIS

The reader should note well that implicit synthesis is not yet ready for use by inexperienced or unwary computing engineers. There are many pitfalls, which are described in the following paragraphs.

PITFALLS OF STABILITY

1. Implicit computation with algebraic equations is known to be beset with pitfalls; implicit synthesis circuits are even more difficult to stabilize because the problem is dynamic rather than static.

2. When a least-squares circuit is interconnected with implicit synthesis circuits, the likelihood of stability problems is still greater.

3. Each implicit synthesis circuit adds its own transients which must be damped out quickly by means of additions of components to the circuit, such as capacitors and potentiometers.

4. Initial conditions must be selected with some care, to insure that the circuit will be able to proceed to the solution in a stable manner.

5. There are restrictions on the type of excitation function $F(t)$ which can be used, if instability is to be avoided.

6. The settling time of implicit synthesis circuits is variable and may be excessively long if too low a gain G is used.

7. In a digital computer or digital differential analyzer, the time increment Δt, used in conjunction with a given gain G, has an upper limit which must not be exceeded for stable operation.

8. An operational analog circuit has an upper limit on loop gain which cannot be exceeded without incurring undesirable transients or unstable operation.

PITFALLS OF ACCURACY

1. Noise in input signals $x(t)$ and $F(t)$ from tapes or transducers causes the answers to jump erratically, because implicit synthesis always involves one or more stages of differentiation.

2. If noise in the input signals is filtered, then accuracy is lost because of phase shift.

3. If the gain required for stability is too low, accuracy suffers.

4. In some implicit synthesis circuits, there is a steady-state error which increases as the dependent variable approaches zero.

5. In some implicit synthesis circuits under some conditions the settling time is so long that the transient gives an appreciable error.

6. $F(t)$ and $x(t)$ must be measured continually during a run and reported to the computer as a continuously varying voltage; if the inputs are known only at discrete times, the answers will be in error.

7. Scaling of an implicit synthesis circuit is critical if maximum accuracy is desired; it might be necessary in many instances to rescale several times.

8. Nonlinearity of tapes or transducers introduces error into the answer.

9. In a second-order system, the damping B cannot be determined accurately unless M and K are known with great accuracy; fortunately either M or K can be found accurately even if B is not known accurately.

The foregoing list of pitfalls is undoubtedly incomplete. Nevertheless, it should serve to arm computing engineers with a respect for the difficulties confronting development of any implicit synthesis circuit.

On the other hand, once a circuit has been found which works for a particular class of applications, and once the ground rules for its use have been set down on paper, implicit synthesis should be a reliable technique.

Dynamic Least Squares

The method of "dynamic least squares"[2] requires: 1. an implicit synthesis or differencing circuit, for determining the derivatives of the dependent variable, and 2. a least-squares circuit, for finding the characteristics. A linear second-order system will be used as an illustration of the method. One is given $x(t)$ from a system obeying the equation

$$M\ddot{x}+B\dot{x}+Kx=F(t) \tag{23}$$

in which M, B, and K are the characteristics to be determined. Assume for the moment that M, B, and K are known. Then $x_c(t)$ and its derivatives can be determined from an implicit synthesis circuit which compares $x_c(t)$ with $x(t)$, multiplies the difference by a large gain, and treats the result as a fictitious force which is used to correct $\ddot{x}_c$ directly and then indirectly $\dot{x}_c$ and x_c. After the implicit circuit transients die out, x_c and x should be nearly equal, the fictitious force should be nearly zero,and so x_c and its derivatives should be nearly correct. Accordingly, consider now that x_c and its derivatives are known. Then M, B, and K can be determined by least squares, so that the function $M\ddot{x}_c+B\dot{x}_c+Kx_c$ fits the given function $F(t)$ with minimum integrated error squared, i.e., one requires the time integral of ϵ^2 to be a minimum with respect to M, B, and K separately, where $\epsilon=M\ddot{x}_c+B\dot{x}_c+Kx_c-F(t)$. Then, by partial differentiation with respect to M, B, and K separately, and equating to zero, one obtains the following "normal equations" for determining M, B, and K:

Table I. Least-Squares Values

Time, Seconds	No Noise	1%	2%	5%	10%	20%
			Values of C_1: Should Be 0.14285714			
0						
0.05	0.14285715	0.14285716	0.14285715	0.14285715	0.14285715	0.14285715
0.10	85717	85719	85717	85719	85718	85718
0.15	85736	85743	85747	85735	85734	85749
0.20	85806	85797	85772	85813	85775	85773
0.25	85818	85853	85853	85853	85840	85831
0.30	85924	85877	85970	85970	85879	85934
0.35	86300	86299	86206	86229	86337	86338
0.40	87141	87060	87044	86969	86988	87008
0.45	88422	88447	88118	88558	88213	88547
0.50	91515	91437	91067	90946	91668	91659
			Values of C_2: Should Be 1.00000000			
0.05	0.99999935	0.99999917	0.99999902	0.99999959	0.99999930	0.99999948
0.10	9920	9931	9960	9960	9920	9893
0.15	9900	9860	9820	9820	9839	9837
0.20	9718	9718	9759	9716	9757	9762
0.25	9757	9676	9757	9675	9680	9755
0.30	9756	9757	9674	9674	9669	9658
0.35	9508	9426	9426	9590	9424	9411
0.40	9174	9091	9259	9171	9100	9252
0.45	8836	8752	8833	8745	8822	8765
0.50	8409	8324	8239	8335	8162	8198

$$ {}^t_0\ddot{x}_c^2dt\ M+\int_0^t\ddot{x}_c\dot{x}_cdt\ B+ \int_0^t\ddot{x}_cx_cdt\ K=\int_0^t\ddot{x}_cF(t)dt \quad (24)$$

$$\int_0^t\dot{x}_c\ddot{x}_cdt\ M+\int_0^t\dot{x}_c^2dt\ B+ \int_0^t\dot{x}_cx_cdt\ K=\int_0^t\dot{x}F(t)dt \quad (25)$$

$$\int_0^tx_c\ddot{x}_cdt\ M+\int_0^tx_c\dot{x}_cdt\ B+ \int_0^tx_c^2dt\ K=\int_0^tx_cF(t)dt \quad (26)$$

If the characteristics are variable rather than constant, a digital computer can conduct a least-squares analysis on a running set of, say, ten points. In this way the characteristics can be determined in every time interval Δt after the first few. Then, they can be plotted against various possible arguments until each characteristic is a known function of a known argument. The same result can be achieved on an analog computer.

The foregoing procedure has been carried out in the simple case of a first-order system excited by a step function. The system obeys the differential equation

$$C_1\dot{x}+C_2x=F(t) \quad (27)$$

where C_1 and C_2 are constants, x is the dependent variable, and $F(t)$ is the forcing function (in this case, a unit step). This equation was set up in an International Business Machines Corporation 704, using the DIDAS (digital differential analyzer simulator) method for simulating a digital differential analyzer, and taking $C_1=1/7$ and $C_2=1$. Then, the $x(t)$ thus generated, together with $F(t)$, was subjected to analysis by dynamic least squares, which was also programmed by means of DIDAS. The Δt for the calculations was 0.001 second. After every 0.050 second, the results of the dynamic least squares determination of C_1 and C_2 were printed out and a new determination was started. Initially, both x and $\dot{x}$ were zero. After 0.500 second, the calculation was stopped. Similar runs were made with random white noise entering the integrals in the normal equations. The noise was limited to 1%, 2%, 5%, 10%, and 20% of the signal in the five runs made with noise. The values of C_1 and C_2 resulting from these least squares determinations were as presented in Table I.

From Table I, one notes at a glance that at least four significant digits were obtained correctly in each determination, even with 20% noise in the signals to the dynamic least-squares circuit. Thus the method of dynamic least squares can tolerate a large amount of noise without an appreciable degradation of accuracy in results.

A slight degradation may be noted in successive determinations. This is caused by the fact that at later times x is very close to unity and $\dot{x}$ is very close to zero, so that 50 values convey less information than did 50 values at earlier times when x and $\dot{x}$ were changing more rapidly. Moreover, a cumulative error arose in the generation of x.

One reason for the good accuracy in the foregoing results in that 50 points were included in each determination, which made most of the noise cancel itself out. The effect of including fewer points in each least-squares determination may be seen from the results in Table II, in which the noise was limited to 5%. The time increments were 0.001 second in all cases.

It is remarkable that even with only five points per least-squares determination, the results were correct to at least four significant digits, in spite of the 5% noise which was present.

Another series of runs was made with a sinusoidal $F(t)$ instead of a unit step. Again the time interval was 0.001 second. There were 50 points per least-squares determination. In the case of a period of 0.05 second, all values of C_1 were either 0.14285714 or 0.14285715, even for 20% noise. The values of C_2 ranged between 0.99999997 and 1.0000029.

It is believed that the foregoing results establish the validity and practical utility of the method of dynamic least squares for determining the constants in the differential equations representing dynamic systems.

Some Applications of Implicit Synthesis and Dynamic Least Squares

The analytical scientist and engineer are often faced with the problem of obtaining the characteristics in the differential equations representing a complicated system. In many instances there is no known satisfactory way to find one or more of the characteristics, chiefly because of the difficulty of analyzing the results of experiments in which more than one quantity is varied. Nevertheless, these scientists and engineers do the best they can, which might in some instances be a very crude approximation to the true characteristic. Not much is said publicly about such temporary failures of the experimental method, except when it is necessary to blame deficiencies of simulation upon the inadequacy of the equations and input data. Any simulation is only as good as the system equations and characteristics that one is given to work with.

The following offhand list is indicative of the enormous variety of applications awaiting the practical development of implicit synthesis and dynamic least-squares techniques:

AIRCRAFT AND AUTOMOTIVE INDUSTRIES

1. Determination of aerodynamic coefficients from flight tests involving complicated maneuvers; from dynamic tests in wind tunnels with elastic mountings; or from rocket sled tests.
2. Determination of the flutter speed of an airplane or missile surface from subcritical tests of airplanes in flight, wind tunnel flutter models, or rocket sled models, by determining the damping of the flutter mode as a function of frequency.
3. Determination of the flexibility characteristics of any structure.

4. Determination of servo internal characteristics and output impedance.

5. Determination of the characteristics of the interior ballistics of guns, rocket catapults, and the like.

6. Determination of the dynamic characteristics of tires.

7. Determination of the nonlinear dynamic characteristics of airplane control sticks, rudder pedals, and steering wheel linkages.

8. Determination of the dynamic characteristics of landing gears and automotive suspension systems.

9. Determination of the dynamic characteristics of turbojet, turboprop, reciprocating, free-piston, and all other engines, compressors, etc.

10. Determination of the parameters governing transient air flow through a porous material such as a fabric.

11. Determination of the course and speed of a target as an aid to tracking or fire control.

12. Determination of the characteristics of a "satellite" and its upper-atmosphere environment, with knowledge of the history of its motion (i.e., its orbit and perturbations).

13. Design of aircraft landing control systems, mid-air refueling systems, and other special purpose computers using prediction.

PROCESS INDUSTRIES: CHEMICALS, FUELS, MATERIALS, ETC.

1. Determination of the chemical kinetic and other parameters which determine reaction rates in batch or continuous processes.

2. Determination of: 1. the half-lives of the constituents, or 2. the composition, of a radiochemical mixture.

3. Determination of the dynamic rheological characteristics of materials undergoing deformation or flow.

4. Determination of the lumped parameters in sets of ordinary differential equations used to approximate distributed-parameter systems such as heat exchangers.

5. Determination of the characteristics of combustion processes.

6. Determination of neutron process parameters in nuclear reactors.

7. Determination of the dynamic parameters of instruments, transducers, and process control system components.

MISCELLANEOUS

1. Determination of dynamic parameters in biological population systems.

2. Determination of the characteristics of systems for which forecasts are to be made, such as dynamic economic systems (sales of a product, growth of a new industry, etc.).

3. Determination of the dynamic characteristics of a human operator of various devices when he is a part of a loop.

4. Determination of the dynamic characteristics of the regulatory systems in human or animal physiology.

Conclusions

Clearly, if an appreciable number of the foregoing possible applications do materialize, implicit synthesis and dynamic least squares will stimulate a major advance in the art of simulation. One will no longer be so bound by crude linearizations, steady-state representations, neglect of higher order derivatives, use of "eyeball" or "ballpark" values, and such approximations. One will be able to simulate some systems which have never been simulated before, such as complete dynamic human physiology.

If implicit synthesis and dynamic least squares become a practical reality, one may expect the use of transfer functions in systems engineering to yield its throne to a greater extent to the use of differential equations for the representation of nonlinear servos and other nonlinear mechanism in systems engineering.

Table II. Least-Squares Values

Determination	No. of Points 50	20	10	5
	Values of C_1: Should Be 0.14285714			
1	0.14285715	0.14285715	0.14285175	0.14285714
2	85719	85719	85175	85717
3	85735	85736	85190	85726
4	85813	85741	85183	85762
5	85853	85781	85241	85748
6	85970	85869	85199	85880
7	86229	85874	85318	85883
8	86969	85947	85304	85901
9	88558	86104	85492	85897
10	90946	86423	85487	86102
	Values of C_2: Should Be 1.00000000			
1	0.99999959	0.99999783	0.99999925	0.99999842
2	9960	9738	9788	9554
3	9820	9796	9639	8855
4	9716	9466	9231	6427
5	9675	9136	9042	6886
6	9674	9059	9452	4081
7	9590	8509	8721	5837
8	9171	8697	6173	7090
9	8745	8091	6599	2951
10	8335	8249	6813	6608

Even if the applications of implicit synthesis and dynamic least squares are not so widespread as envisioned in the foregoing, one may expect them to form a new chapter in future books on computation.

There might well be important implications for the future of experimental science and technology, quite apart from providing imput data for simulation. To appreciate this possibility, one may look back upon the history of experimentation.

Consider first the field of static experiments, in which practical men seek a better process or material by varying factors to find the best combination. Traditionally this was done in accordance with the philosophy of "varying one thing at a time." because otherwise no one knew of any way to analyze the results. During the past half-century, thanks to the efforts of statisticians, methods have been developed for interpreting experiments in which many factors have been varied at once. In fact, the development of these statistical methods of analysis has permitted an entirely new philosophy of experiment design to spring into use. The modern theory of the design of experiments permits great economies on the number of experiments required by providing for the deliberate variation of many factors at once. As a dividend, the resulting information is far more accurate and broad in scope of validity.

In the same way, in the field of dynamic experiments it has been necessary to limit the generality of experiments, because techniques have not been available for analyzing the results of a general dynamic experiment. It is customary to set certain variables on zero or other standard values, restrict excitation functions to sine waves, unit steps, or ramps, and otherwise specialize each experiment, in order to permit data reduction. It is perhaps not too farfetched to conceive of implicit synthesis and dynamic least squares being commonly used to analyze the results of dynamic experiments and to display the desired characteristics of the system under study. If this takes place, it will lead to a minor revolution in the philosophy of dynamic experimentation, because it will relax restrictions on what can be varied and, indeed, will encourage deliberate variation, in order to investigate effects which do not appear under restricted conditions.

The ultimate in the application of implicit synthesis and dynamic least squares would be to conduct a single experimental run on an unknown system and to display as a result each characteristic of the system. The other extreme would be the conventional use of many experimental runs for the determination of each system characteristic, using paper-and-pencil or classical digital-computer methods of fitting discrete points for data reduction. In between these two extremes lies an interesting exercise in operations research. If the chief cost is from experimental time, then it would pay to invest a little money in computation in order to save experimental costs. This is clearly desirable in the case of jet engine testing, airplane or missile flight testing, rocket sled testing, nonproduction testing of a chemical plant, and so on. On the other hand, if the cost of making experiments is small compared with the cost of using a computer, then implicit synthesis and dynamic least squares would not be economical, though possibly still desirable.

All in all, the future possibilities of implicit synthesis and dynamic least squares are intriguing, even though the road has pitfalls. Surely the probability of success in at least a few applications, multiplied by the eventual payoff, is high enough to justify some research and development on implicit synthesis and dynamic least squares by nearly every organization which uses computers for the study of dynamic systems.

Nomenclature

A = amplitude of excitation in Duffing's equation
B = damping coefficient
C_1 and C_2 = coefficients to be determined
$F(t)$ = forcing function or excitation function
G = a large positive number
K = spring coefficient
M = mass coefficient
P = potentiometer setting
t = independent variable
x = dependent variable
y = refer to Fig. 9
α = coefficient of linear term in Duffing's equation
β = coefficient of cubic term in Duffing's equation
Δt = time increment in integration on digital computer
ϵ = error
λ = unknown constant to be determined
ω = frequency of excitation in Duffing's equation
τ = time constant

Subscript c means "computed value of"

References

1. Electronic Instruments (book), I. A. Greenwood, Jr., J. V. Holdam, Jr., D. Macrae, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, pp. 15–21.

2. Data Reduction for the Free Flight Spark Ranges, C. H. Murphy. *Report no. 900*, Ballistic Research Laboratory, Aberdeen, Md., 1954, p. 16.

The ON/K Carrier System

G. T. MASTERS
ASSOCIATE MEMBER AIEE

DUE to the recent growth of telephone requirements in northern California, a constant search is being made for better methods of using the basic telephone facilities in this area. One of the facilities which was looked at was the cable pairs used by the *K* carrier systems. This system, by design, is a long-haul carrier. That is, it is designed to provide circuits that have transmission characteristics suitable for 4,000-mile circuits. However, since the capacity of this system is only 12 circuits it was seen to be advantageous to devise some method to increase the capacity of the cable pairs. The circuits derived by such an addition could be short-haul, of 100 to 200 miles. The requirements for new long-haul circuits are being filled by the *L1*, *L3*, *TD2*, and in the future, *TH* carrier systems. The objective then of examining the *K* carrier cable pairs was to provide a method of deriving additional short-haul voice channels on these same pairs. One of the stipulations for any method of accomplishing this, however, is that no degradation of the *K* carrier systems should result from the modification or addition. A second stipulation, since the need for the additional channels was fairly urgent, is that any instrumentality employed should use as much existing equipment as possible in order to minimize the development time. With these factors in mind, a study was started which eventually led to the development of the *ON/K* system. This paper describes the development and the use of these systems in the northern California area of the Pacific Telephone Company.

The Development

Previous to this time several of the Bell System companies had stacked type *H* and other types of carrier systems under the *K* systems to provide single-channel service to points along the main *K* carrier route. However, the stacking of any carrier above the *K* system had not been attempted before this time. The Bell Telephone Laboratories became interested in this problem in response to a suggestion from the Northwestern Bell Company. An announcement was made to the operating companies about "putting the squeeze on the *K* cables," and a survey was made to see which of the operating companies would have application for such an arrangement. The results of the survey indicated that development of the system would be beneficial to several of the companies. The Northwestern Bell Company had an immediate need, and the Pacific Company would also require the system at an early date. These two companies agreed to loan engineering personnel to the laboratories during the development and field trials of the system.

The *K* carrier system has been described in previous papers.[1,2] A recapitulation of some of the basic characteristics is in order, to point up some of the problems encountered in the development of the *ON/K* system. The *K* system operates in the frequency range of 12 to 60 kc, using two cables, one in each direction of transmission. On cables of 19 gage, the repeaters are nominally spaced at a distance of 17 miles. It is necessary to provide special balancing arrangements to equalize the coupling between the pairs used for this carrier in order to minimize the crosstalk coupling. On the original *K* carrier systems, single-section balancing was employed, i.e., balancing equipment was provided at each repeater point. Later,

---

Paper 58-1005, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted May 12, 1958; made available for printing July 10, 1958.

G. T. Masters is with the Pacific Telephone and Telegraph Company, San Francisco, Calif.

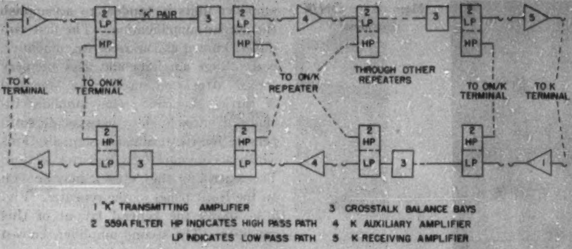


Fig. 1. Typical K carrier system equipped with 559A filters

it was found that several repeater sections could be balanced at a single point. This scheme is known as multisection balancing. Although single-section balancing provides the best over-all crosstalk suppression and is the arrangement used on most routes, multisection balancing offers certain economies and has been used on a few routes. A complete discussion of the methods of treatment for crosstalk is given in cited articles.

The first thing necessary to add channels to the *K* carrier cable pairs is to increase the frequency bandwidth of the repeaters. The present *K* repeaters have a bandpass of 12 to 60 kc. With the addition of 12 channels of *ON* it is necessary that the repeaters handle a bandwidth from 12 to 136 kc. One of the methods suggested to accomplish this was to modify a standard *K* repeater. Another was to use a *J* carrier repeater. These ideas proved impractical and were rejected. This then meant the development of a new repeater. The development objectives were kept in mind, and it was decided to retain the *K* carrier system in essentially its present form, and to develop the new repeater to pass only that portion of the band that was being added. This means that it is necessary to "split" the frequency band not only at the terminals but at each repeater point. A suitable filter to split the frequency band then became the next item for development. The Western Electric *559A* filter was designed for this purpose.

It can be shown that the addition of these filters in a crosstalk balancing section would upset the balance. Therefore, it was felt that this system should only be used in single-section balanced cables. Fig. 1 shows how these filters are placed in the line. The insertion of the filters on the line deteriorates the equalization of the *K* line, making it necessary to re-equalize the line. Previously, the standard method of equilization of *K* carrier lines utilized the incoming level of the pilots and test tones applied at the transmitting *K* terminal. This method was time-consuming, and had the additional disadvantage that the person making the equalization could not see the characteristics of the complete frequency spectrum of the *K* line at one time. This was because the receiving test equipment used during the test is capable of reading only one frequency at a time. Due to the amount of re-equalization to be done at present and in the near future, a new method of re-equalization was developed. The method which meets many of the requirements for re-equalizing *K* lines utilizes a spectrum analyzer, a pulse generator, a level monitoring meter, and a specially equipped portable *K* line amplifier.

The spectrum analyzer is a commercial test set made by Panoramic Radio Products, Inc. It can be adjusted to observe any portion of the spectrum up to 300 kc. Any signals present in the particular band chosen appear on a cathode-ray tube as vertical pips located horizontally in the order of their frequency. The amplitude of the individual pip is a representation of the power at that frequency. A picture of the unit is shown in Fig. 2.

The pulse generator is a 4-kc square-wave generator designed by the Bell Telephone Laboratories. Since the output of the generator is a square wave, it is rich in harmonic content. These harmonics are used for the re-equalization tests. When the generator is connected to a *K* line, the fundamental is below the band of the system, and the harmonics, one every 4 kc, are transmitted to the receiving end of the system. The harmonic output of the generator is essentially constant in the passband of the *K* carrier so that any variation in level that is noted at the receiving end is due to the characteristics of the line.

The level monitor meter is a simple thermocouple meter designed by the laboratories and is used to provide a constant check of the over-all line level during re-equalization.

To provide a rapid method for substituting all of the available equalizers into a *K* line being re-equalized, a portable *K*-line amplifier was equipped with all the possible equalizers available on a switched basis. Fig. 3 shows the complete amplifier and equalizer. In addi-

Fig. 2. Panoramic spectrum analyzer

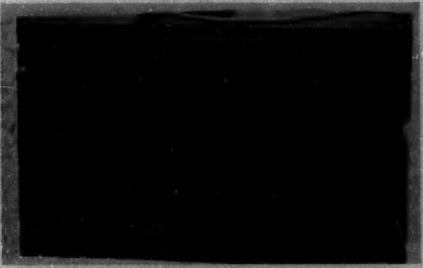


Fig. 3. Amplifier-equalizer equipment

Fig. 4. ON/K terminal

tion, a photographic method of recording data was developed which simplified both the data-taking and analyzation.

The pulse generator is placed at the transmitting *K* terminal on the section to be re-equalized. The special *K* amplifier is patched in place of the regular *K* amplifier and the equalizer controls adjusted until the proper equalization is observed on the spectrum analyzer connected to the output of the amplifier. The strapping on the regular amplifier is then changed to agree with the settings obtained on the test amplifier. The regular amplifier is then placed back into the line, and a check made with the spectrum analyzer to verify the equalization. Photographs of the spectrum analyzer presentation may be taken at any time to be used for comparison or record purposes.

The Terminal

The *ON/K* system in its present form uses a portion of the standard *ON* terminal. The *ON* carrier system has been covered in a previous paper;[3] therefore, no detailed description will be given here. Fig. 4 shows a typical installation of the *ON* terminal equipment as used for the *ON/K* system. Due to the frequency restriction imposed by the available band, it is impossible to use the complete *ON* terminal, and presently only three groups are used for this application. This arrangement gives a gain of 12 channels. In the near future the *ON* terminal is to be modified. It presently has five 4-channel groups but the change will result in six 4-channel groups. At that time, 16 channels may be obtained when using this terminal equipment for *ON/K*. Arrangements are available to extend the *ON/K* systems over open wire or *N* carrier cable pairs by the use of *ON* or *O* junction equipment. When the modified *ON* terminal is used, the ability to interconnect with *O* carrier is lost, so this system can be extended only over cable facilities.

The Repeater

The repeaters used in the *ON/K* system are one of the most interesting features of the system. It was not practical to interset repeaters in the *K* repeater section to amplify the *ON/K* band because of the crosstalk balancing arrangement. It becomes necessary to design a repeater that will properly amplify the *ON/K* signals for transmission over the complete *K* repeater section. The loss of a standard *K* repeater section at the frequency of operation (80 to 136 kc) is approximately 100 decibels. It was decided to use two amplifier units in tandem to accomplish the needed amplification. The first amplifier, known as the receiving amplifier, is a voltage amplifier and uses circuitry copied after that used in a standard *N* carrier repeater. This amplifier includes a variable slope network to compensate for the residual misequalizations which accumulate in a repeater section. The control for the network may be seen on the front panel of the repeater. Fig. 5(A) shows the general layout of this amplifier. The second amplifier, known as the transmitting amplifier, is a power amplifier. This unit has an output of +20 decibels referred to 1 milliwatt, and is self-regulating to keep the output constant. The regulation is accomplished by a thermistor-controlled network in the feedback path of the amplifier. Fig. 5(B) shows the layout of this unit.

The packaging of these amplifiers was another example of the development-saving methods that were used on this system. The die-cast stampings that are in use for the twin channel modules of the *O* carrier were used for these amplifiers with only minor changes. This meant that no new dies or other fabrication machinery were needed to produce the amplifier or power module mountings. Fig. 5(C) shows one of the repeater mountings with the amplifiers and power module in place.

The filament supply for these amplifiers is alternating current, as in the standard *K2* repeater. A transformer that is mounted at the top of the equipment bay is used to change the 60 volts used for office distribution to 19 and 6 volts for the filaments. The 60 volts used for office distribution are normally obtained by the use of a transformer from standard commercial power sources. In the event of a commercial power failure, the filament supply is switched from the commercial mains to an emergency vibrator supply which operates from the office battery. The plate supply for both of the amplifiers is obtained from the office batteries. At a *K* auxiliary station the plate supply is 152 volts. At *K* main stations the common plate battery is 130 volts. This voltage is adequate for the receiving amplifiers but the transmitting amplifiers need the higher voltage in order to provide the output power required. To obtain 152 volts at *K* main stations, it was necessary to put the negative 24-volt supply in series with the 130-volt supply. A terminal block with movable links is provided in the power module to make the necessary voltage connections for either auxiliary or main station use. The power

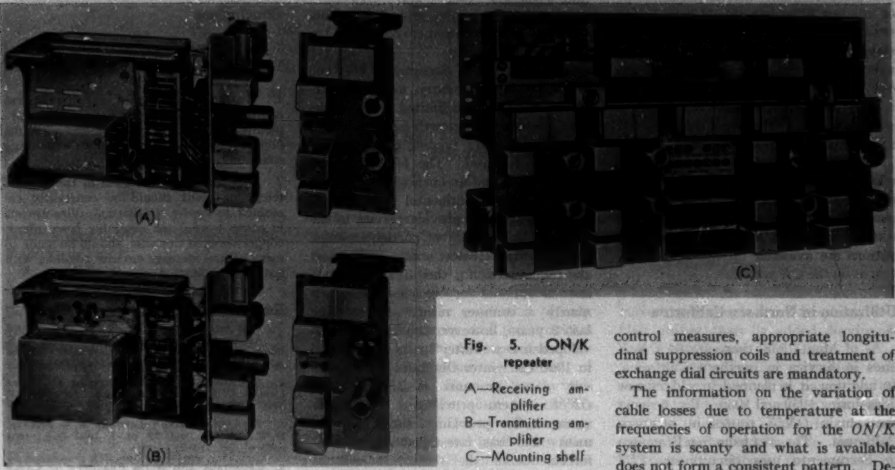


Fig. 5. ON/K repeater

A—Receiving amplifier
B—Transmitting amplifier
C—Mounting shelf

module also contains the necessary fusing and alarm circuits for the repeater mounting. It may be seen in the pictures of the unit that pin-type jacks are provided for the in-service measurement of all voltages used by the repeater. It will also be noted that there are four pin jacks on the front of each amplifier. These are used to make in-service tests of the condition of tubes in the amplifiers.

On repeater sections shorter than the nominal distance, it is necessary to use some networks to build out the section electrically to the length of a standard repeater section. These networks are normally placed between the receiving and transmitting amplifier. This gives an advantage in signal-to-noise ratios. On very short sections where large amounts of build-out are required, a portion of the network can be placed before the receiving amplifier to prevent overloading. The networks used in this application are the same as those that were designed for the *N* carrier system. The space for mounting an adequate number of networks to care for most applications is provided in the repeater mounting unit. This space can be seen between the terminal strips in Fig. 5(C). On short repeater sections an auxiliary panel is supplied to mount the additional networks.

Transmission Characteristics

The over-all transmission characteristics of the system are comparable to those of the *N* or *ON* carrier systems. The channel noise and crosstalk objective is 26 decibels adjusted at the zero level point on systems up to 200 miles. Experience has proved that these values have been attained or bettered on the *ON/K* systems now operating. Due to the low line levels encountered, down to −80 decibels referred to 1 milliwatt, extreme care is necessary in the wiring layouts. Special shielding practices have been developed and the placement of the office wiring is carefully specified in the installation procedures. The *K* cable pairs must be examined for proper noise control measures, appropriate longitudinal suppression coils and treatment of exchange dial circuits are mandatory.

The information on the variation of cable losses due to temperature at the frequencies of operation for the *ON/K* system is scanty and what is available does not form a consistent pattern. Design calculations indicated that the regulation deviations encountered in the *ON/K* system would be similar in magnitude to those in a normal *ON* system on distances up to 200 miles. Experience has shown that circuits obtained from this system are entirely adequate in this respect for short-haul usage. For those systems that do not perform properly or for systems in excess of 200 miles in length an adjustable deviation equalizer has been provided. With the use of this equalizer several systems of lengths up to 400 miles have been planned.

Because of the misequalization caused by the addition of the *559A* filters in the *K* system lines, no more than 500 miles

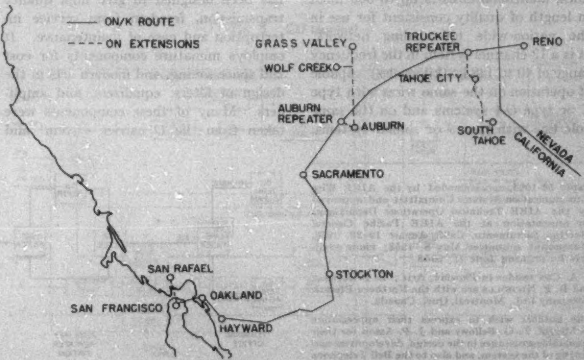


Fig. 6. ON/K routes in northern California

of any one *K* system should have *ON/K* systems operating above it. This restriction will not ordinarily hamper the operation of the *ON/K* system in any one Bell System operating company, but it will mean that careful co-ordination will be necessary between companies to see that this limit is not exceeded.

The gain frequency characteristics, mainly controlled by the channel terminal equipment, are the same as those observed on a normal *ON* system. Of course, the same levels and signalling options are available on the *ON/K* system as on the *ON* system.

Utilization in Northern California

As of May, 1958, there are 64 *ON/K* lines in service in northern California. In addition, it is planned to add 21 new lines during 1959 and 1960. Fig. 6 shows the main *ON/K* routes in northern California. The area from San Francisco to Sacramento uses circuits on *TD2* radio, *N* carrier, *ON* carrier, and *ON/K* carrier to meet the requirements. The use of *ON/K* carrier for a portion of these requirements has allowed a deferment of other and more expensive types of route reinforcement and has provided another means of diversification of routing for the protection of service.

The area east of Sacramento to Reno is mountainous country and the telephone facilities are limited. There has been considerable growth in this area in the last few years, tending to exhaust all available facilities. The few voice frequency pairs and an open-wire lead, part of the old transcontinental wire, are the basic facilities to provide service to the area. A community of particular interest is the Tahoe resort area. In the past the requirement for this area was above normal but was very seasonal; it is primarily a summer resort. During the last 2 years, however, this area has developed into a winter sports center, and in 1960 the winter Olympics will require very large amounts of service. The *ON/K* system provides a means for economically providing these requirements and has forestalled costly reinforcements of the route. It will be noted that all of the communities of interest served by this route are not on the main cable; therefore, many of the *ON/K* systems are extended to these points by means of *ON* carrier.

Conclusions

The experience gained during the use of the *ON/K* systems has proved again that:

1. By judicious use of existing instrumentalities the necessity for new development can often be minimized, and all of the objectives still be gained, in an expanding communication network.

2. The basic facilities used in the present telephone plant should be constantly examined to assure maximum utilization or, in other words, the available basic plant, copper wires in this case, should be made to carry all the message circuits possible, with economic limitations always in mind.

References

1. An Improved Cable Carrier System, **H. S. Black, F. A. Brooks, A. J. Wier, I. G. Wilson.** *AIEE Transactions*, vol. 66, 1947, pp. 741–46.

2. A Carrier Telephone System for Toll Cables, **C. W. Green, E. I. Green.** *Ibid.*, vol. 57, May 1938, pp. 227–36.

3. Type ON Carrier Telephone, **R. D. Fracassi, H. Kahl.** *Ibid.*, vol. 72, pt. I, 1953 (Jan., 1954 section), pp. 713–21.

4. Crosstalk and Noise Features of Cable Carrier Telephone, **M. A. Weaver, R. S. Tucker, P. S. Darnell.** *Bell System Technical Journal*, New York, N. Y., vol. 17, 1938, pp. 137–61.

5. Cable Crosstalk—Effect of Non-Uniform Current Distribution in the Wires, **R. N. Hunter, R. P. Booth.** *Ibid.*, vol. 14, 1935, pp. 179–94.

Type OJ Carrier Telephone System

J. A. COY
MEMBER AIEE

E. S. KELSEY
MEMBER AIEE

B. P. NICHOLLS
ASSOCIATE MEMBER AIEE

THE OJ carrier system is designed to provide economically, on open-wire lines, additional circuits up to 500 miles in length of quality consistent for use in the nation-wide toll dialing network. It is a 12-channel system in the frequency range of 40 to 148 kc (kilocycles), capable of operation on the same wires with type *C* or type *OA* systems and on the same pole line with type *J* or similar systems. It is intended for the same field of use which type *J* has filled successfully and has been designed to give high quality transmission, freedom from service interruption and ease of maintenance. It employs miniature components for cost and space saving, and modern arts in the design of filters, equalizers, and amplifiers. Many of these components were taken from the *O* carrier system[1] and hence the designation *OJ* for a system similar to *O* in its composition and occupying the type *J* field.[2]

The longer and more important open-wire lines usually have *J* carrier systems on some of their wires. Type *O* systems, employing a frequency allocation not compatible with *J*, cannot be added to such lines for any considerable distance. *OJ* however can be added and can also be extended out over branch lines or can readily be connected to *O* systems on such branches or to type *ON* carrier[3] in cable extensions. When so desired, circuits can be dropped or branched in groups of four at any repeater point.

A typical *OJ* carrier application on a

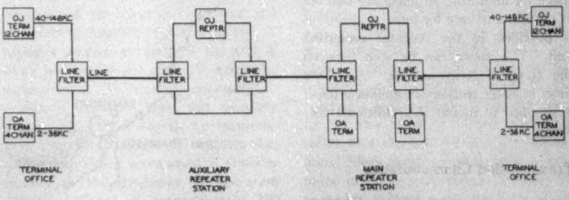


Fig. 1. Typical OJ and OA application

Paper 58-1004, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted May 8, 1958; made available for printing June 12, 1958.

J. A. Coy resides in Phoenix, Ariz.; E. S. Kelsey and B. P. Nicholls are with the Northern Electric Company Ltd., Montreal, Que., Canada.

The authors wish to express their appreciation to Messrs. T. G. Fellows and J. P. Aston for their invaluable assistance in the design, development and testing of the system, and also to the Bell Telephone Company of Canada for the use of their facilities and for the assistance of their personnel during the field trial.

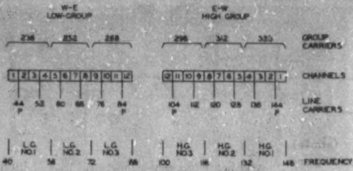

W-E
LOW-GROUP
E-W
HIGH GROUP
236
252
268
296
312
328
GROUP CARRIERS
1 2 3 4 5 6 7 8 9 10 11 12
12 11 10 9 8 7 6 5 4 3 2 1
CHANNELS
44 P
52
60
68
76
84 P
104 P
112
120
128
136
144 P
LINE CARRIERS
L.G. NO.1
L.G. NO.2
L.G. NO.3
H.G. NO.3
H.G. NO.2
H.G. NO.1
40
56
72
88
100
116
132
148
FREQUENCY

Fig. 2 (left). Frequency allocation, type OJ carrier system

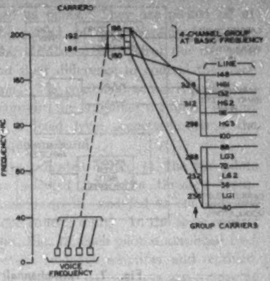

CARRIERS
200
160
120
80
40
0
FREQUENCY-KC
192
184
196
180
4-CHANNEL GROUP AT BASIC FREQUENCY
LINE
148
HG1
132
HG2
116
HG3
100
88
LG3
72
LG2
56
LG1
40
328
312
296
268
252
236
GROUP CARRIERS
VOICE FREQUENCY

Fig. 3 (right). Modulation plan, type OJ system

line in association with an *OA* system to derive a maximum of 16 carrier channels is shown in Fig. 1.

As in *N* and *O* carrier systems, the assembly of the units is designed for a maximum proportion of wiring, testing and adjusting to be done in the factory and with a minimum of field engineering and installation effort. Die-cast frameworks for units and mountings insure mechanical rigidity and good fit between parts.

Frequency Allocation Plan

The frequency allocation of the *OJ* system is shown in Fig. 2. W-E (west-east) transmission in the range 40 to 88 kc and E-W (east-west) transmission in the range 100 to 148 kc co-ordinates with the type *J* allocations. The frequency space below 40 kc is left clear for *C* or *OA* carrier. Twin-channel operation, that is, the transmission of two channels as the sidebands of a common carrier, requires the generation of only one carrier frequency and the provision of only one gain regulator per two channels. The discrimination required of the channel band filters is less by this method of operation since the carriers are 8 kc apart and the unwanted energy that gets through the band filter from some other twin will demodulate as frequencies above 4 kc. These can be removed by a simple voice frequency roof filter rather than by added discrimination in the channel band filter.

By taking advantage of built-in compandors, as described in papers on the type N^4 and type *O* carrier systems, it has been possible to reduce the side-band energy transmitted to the line very substantially without the system becoming oversensitive to noise and

Fig. 4. OJ carrier terminal equipped for 12 channels

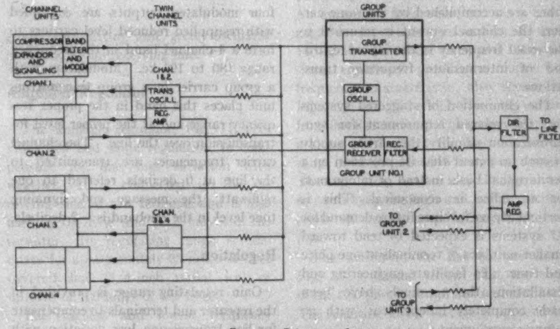

CHANNEL UNITS
TWIN CHANNEL UNITS
GROUP UNITS
COMPRESSOR
EXPANDOR AND SIGNALING
BAND FILTER AND MODEM
CHAN.1
CHAN 1&2
TRANS OSCILL
REG. AMP
CHAN.2
CHAN.3
CHAN 3&4
CHAN.4
GROUP TRANSMITTER
GROUP OSCILL
GROUP RECEIVER
REC. FILTER
GROUP UNIT NO.1
DIR FILTER
TO LINE FILTER
AMP REG
TO GROUP UNIT 2
TO GROUP UNIT 3

Fig. 5. OJ terminal

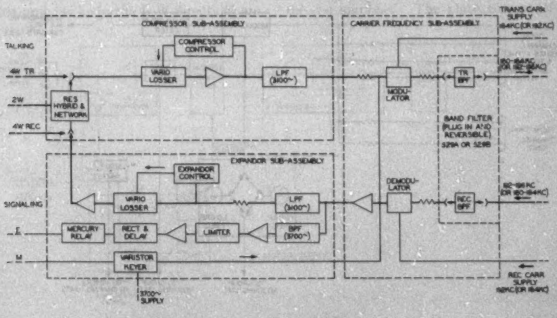

COMPRESSOR SUB-ASSEMBLY
TALKING
4W TR
2W
4W REC
RES HYBRID & NETWORK
COMPRESSOR CONTROL
VARIO LOSSER
LPF (3100~)
EXPANDOR SUB-ASSEMBLY
EXPANDOR CONTROL
SIGNALING
VARIO LOSSER
LPF (3100~)
E
MERCURY RELAY
RECT & DELAY
LIMITER
BPF (3700~)
M
VARISTOR KEYER
3700~ SUPPLY
CARRIER FREQUENCY SUB-ASSEMBLY
TRANS CARR SUPPLY 184KC (OR 192KC)
MODU-LATOR
180-184KC (OR 192-196KC)
TR BPF
BAND FILTER (PLUG IN AND REVERSIBLE) 529A OR 529B
DEMODU-LATOR
REC BPF
192-196KC (OR 180-184KC)
REC CARR SUPPLY 192KC (OR 184KC)

Fig. 6. Channel unit

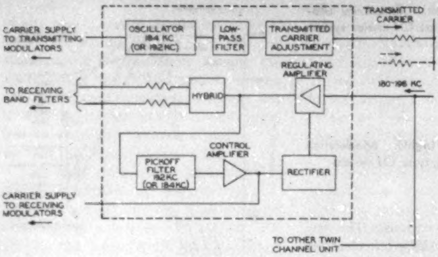

CARRIER SUPPLY TO TRANSMITTING MODULATORS
OSCILLATOR 184 KC (OR 192KC)
LOW-PASS FILTER
TRANSMITTED CARRIER ADJUSTMENT
TRANSMITTED CARRIER
REGULATING AMPLIFIER
TO RECEIVING BAND FILTERS
HYBRID
180-196 KC
CONTROL AMPLIFIER
PICKOFF FILTER 192KC (OR 184KC)
RECTIFIER
CARRIER SUPPLY TO RECEIVING MODULATORS
TO OTHER TWIN CHANNEL UNIT

Fig. 7. Twin-channel unit

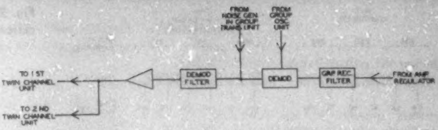

FROM NOISE GEN IN GROUP TRANS UNIT
FROM GROUP OSC UNIT
TO 1 ST TWIN CHANNEL UNIT
TO 2 ND TWIN CHANNEL UNIT
DEMOD FILTER
DEMOD
GRP REC. FILTER
FROM AMP REGULATOR

Fig. 9 (above). Group receiving unit

crosstalk. The compandor advantage also makes it unnecessary to stagger frequency allocations as was done in type *J*, to reduce intersystem crosstalk.

As in the type *O* system the channel carriers are transmitted at a reduced level and employed at the distant terminal for channel regulation and for demodulation. Since the modulation at one terminal and the demodulation at the other are accomplished by the same carrier, the channel output is returned to the exact frequency of the input, regardless of intermediate frequency translations.

The elimination of staggered systems and the relaxed requirement for synchronization simplify the carrier supply to such an extent that its provision on a per terminal basis instead of in common for an office is economical. This is particularly useful since future demand for *OJ* systems is expected to tend toward smaller numbers of terminals at one place and time. To facilitate engineering and installation, the terminals have been made completely independent, with no common equipment.

Modulation Plan

The modulation plan of the *OJ* system is shown in Fig. 3. The 12 channels are arranged in three groups of four channels each. In each of these 4-channel groups, four single side-band carrier channels are generated: one as the upper sideband and one as the lower sideband of a 184-kc carrier; one as the upper and one as the lower sideband of a 192-kc carrier. The four modulated outputs are combined with resupplied reduced level carriers to form a 4-channel band in the frequency range 180 to 196 kc. Modulation with a group carrier in a group transmitting unit places this band in the proper frequency range and at the proper level for transmission over the line. The channel carrier frequencies are transmitted to the line at 0 decibels referred to one milliwatt, the message and signaling tone level in the sideband is −2 decibels.

Regulation

Gain regulating range is provided in the repeater and terminals to compensate for line transmission loss variation with weather from dry to wet and to 1/4-inch radial thickness of ice. Flat and slope regulation are provided at each repeater and in the receiving side of each terminal. The highest and lowest carriers in the W-E band, 84 and 44 kc, and in the E-W band, 144 and 104 kc serve as pilot tones. Flat regulation is by means of a variolosser controlled by the sum of the two pilots; slope regulation, by means of a variable slope equalizer controlled by the difference of the two pilots. Mop-up regulation and equalization is provided in the twin-channel units, controlled by the carrier for each pair of twin channels.

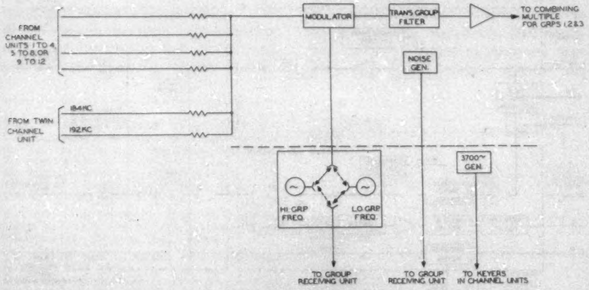

FROM CHANNEL UNITS 1 TO 4, 5 TO 8, OR 9 TO 12
FROM TWIN CHANNEL UNIT
184KC
192KC
MODULATOR
TRANS GROUP FILTER
TO COMBINING MULTIPLE FOR GRPS 1,2 & 3
NOISE GEN.
3700~ GEN.
HI GRP FREQ.
LO GRP FREQ.
TO GROUP RECEIVING UNIT
TO GROUP RECEIVING UNIT
TO KEYERS IN CHANNEL UNITS

Fig. 8. Group transmitting unit and group oscillator unit

Terminal

Fig. 4 shows an *OJ* 12-channel terminal. The principal assemblies are from top to bottom: mounting and units for channels 1 to 4, the same for channels 5 to 8 and for 9 to 12 and finally the amplifier-regulator in the common receiving circuit. Fig. 5 is a block schematic of an *OJ* terminal, divided into functional areas. Each of these areas is also a physical unit or sub-assembly. In the following description, those parts previously described in the paper on the type *O* carrier telephone will be passed over briefly and more attention given to those peculiar to *OJ* carrier.

Each channel unit, Fig. 6, consists of three subassemblies; the compressor, the expandor and signaling circuit, and the carrier frequency circuit. The built-in 3,700-cycle out-of-band signaling for the transmission of dialing and supervisory signals, is optional. This signaling system has been described in the papers on *N* and *O* carrier systems.[1,4] The channel units are identical with those of *O* carrier. Briefly they include the compandor, the signaling circuit, the modulator, demodulator, and associated channel band filters.

A twin-channel unit, Fig. 7, like that of type *O*, includes the oscillator supplying modulators of two channels with 184-kc or 192-kc carrier frequency, the pick-off filters and amplifier supplying the demodulators of the same two channels with carrier frequency taken from the line, and the receiving channel gain regulator.

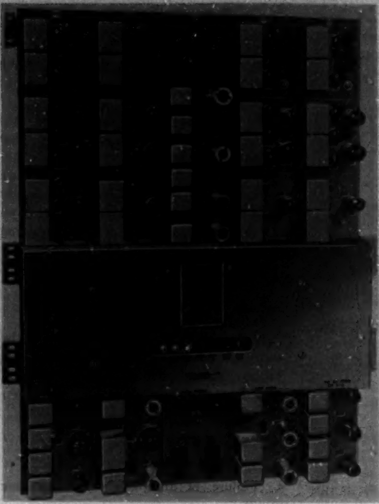


Fig. 10. Four-channel group assembly

Each of the three group transmitting units, Fig. 8, consists of a modulator and associated filter which places a basic 4-channel group in the proper location in the line frequency spectrum, and a 2-stage feedback amplifier, with relatively flat transmission from 40 to 148 kc. This unit also includes a noise generator from which a controlled level of noise is introduced into the receiving side of the system to mask intelligible crosstalk during silent intervals. Each group oscillator unit, also shown in Fig. 8, supplies two carrier frequencies one to the modulator in the group transmitting unit, the other to the demodulator of the group receiving unit. The group oscillator unit also contains a 3,700-cycle oscillator supplying signaling frequency to the four channels of the group.

The group receiving unit, Fig. 9, comprises a group receiving filter which selects one of the three 4-channel groups, the demodulator which translates the group from its line frequency allocation to the basic group frequency band, 180 to 196 kc, a demodulation filter to remove unwanted modulation products and a two stage amplifier. Noise from the noise generator in the group transmitting unit is introduced at the output of the group demodulator to mask intelligible crosstalk as previously mentioned.

Fig. 10 shows the 4-channel group terminal assembly, comprising four channel units, two twin-channel units, the group transmitting unit, group oscillator unit, and group receiving unit.

Amplifier-Regulator

The amplifier-regulator, Fig. 11, provides flat and slope regulation to compensate for line loss variation with weather. It is connected in the receiving side of the terminal and is also used as the E-W and W-E halves of the repeater. Its regulating range is adequate for approximately 33 miles of sleet covered line (1/4-inch radial ice) or upward of 100 miles of wet line.

The highest and lowest frequency channel carriers in each 12-channel band (44 and 84 kc in the W-E direction and 104 and 144 kc in the E-W direction) are used as pilots. They are picked off at the output of the amplifier-regulator, their sum used to control the flat gain and their difference to control the slope gain. The combined action is such as to maintain the amplifier-regulator output at a fixed level, constant across the frequency band.

As indicated in Fig. 11, the two pilots are selected from the through transmission circuit, and fed to the flat and slope control units. In the slope control unit, Fig. 12, each pilot is amplified by a 2-stage feedback amplifier and rectified. The resulting d-c voltages are connected series opposing to the input of a single stage d-c amplifier. The plate current of this amplifier passes through a thermistor which is the controlling resistance element of a variable slope network. A small difference in the two pilot currents controls a relatively large change in the thermistor current and swings the network characteristic from maximum slope in one direction, to maximum slope in the other, as indicated by the variable equalizer characteristics, *A* and *B*, in Fig. 13 (low group) and Fig. 14 (high group). An associated fixed equalizer has a fixed negative slope, as shown in Figs. 13 and 14. The sum of the fixed and variable slope equalizers provide a combined frequency characteristic, also shown in Figs. 13 and 14, which may be varied from flat loss across the frequency band to a slope of 30 db in the high frequency range or 18 db in the low frequency range. The regulating action sets the equalizer so as to keep the high and low frequency pilots equal at the amplifier-regulator output.

The action in the flat control unit, Fig. 15, is similar except that the sum of the two pilots is amplified and rectified. The resulting direct voltage is opposed to a constant direct reference voltage and the difference is impressed on the input of the d-c amplifier. The plate current of

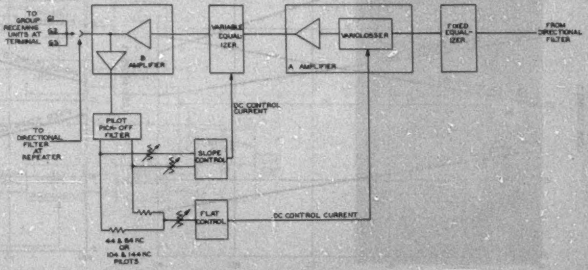


Fig. 11. Amplifier-regulator

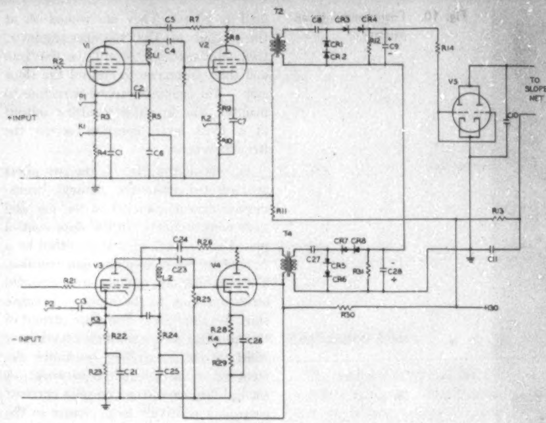


Fig. 12. Circuit of slope control unit

this amplifier passes through a variolosser shunted across the main transmission path. The loss in this unit is thus adjusted so that the total output of the amplifier-regulator remains constant. The bias voltage is stabilized by an avalanche diode.

The variolosser, Fig. 16, consists of a pair of diffused silicon diodes connected opposing in shunt across a relatively high impedance circuit. With no direct current flowing the a-c impedance of this shunt is very high so that negligible loss is introduced. As the direct current increases the a-c impedance falls. The range of direct current employed in the *OJ* unit is approximately from 0.01 to 12 milliamperes with a resultant a-c shunt impedance of 3,000 to 10 ohms and an insertion loss of 10 to 55 db. Interchannel modulation in this circuit has been reduced by the introduction of a capacitor at the midpoint of the output transformer. Without this capacitor the effect of any unbalance between the two diodes is amplified by d-c flow through the shunt path afforded by the transformer winding.

By the combined action of the two control units, the one adjusting the slope to keep the two pilots equal, the other adjusting the flat loss to keep their sum constant, the line loss is automatically built out to give a constant output from the amplifier-regulator to the outgoing line at a repeater or to the group receiving units at a terminal.

Alarm Circuits

Alarm circuits are provided in the terminal mountings to give individual visual indications of the following troubles:

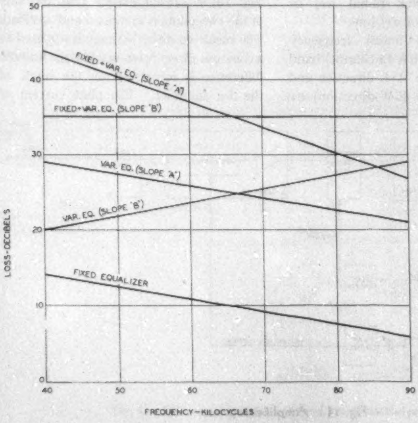
1. Failure of the 48-volt fuse
2. Failure of the 130-volt fuse
3. Loss of incoming carrier
4. Failure of the 3,700-cycle signaling supply

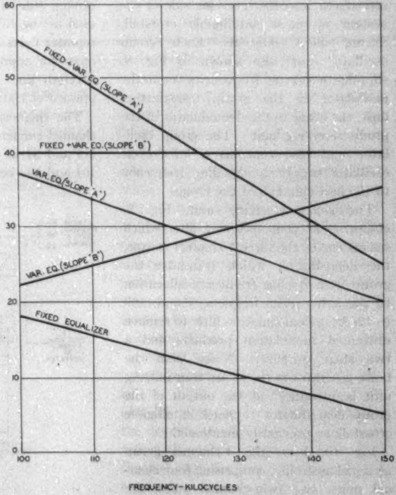
Provision is also made for operating the central office visual and audible alarms. These alarms can be cut off by operation of an "Alarm Release" key. Visual and audible warning will again be given when the system is restored to service.

In the case of 3,700-cycle or carrier failure, the alarm circuit opens the *E*

Fig. 13 (below). Frequency characteristic of low group equalizer

Fig. 14 (right). Frequency characteristic of high group equalizer





leads to the office supervisory or dial equipment, to give called subscribers a chance to disconnect their telephones. After an interval of about 10 seconds, controlled by the operation of a thermal relay, ground is supplied on the *E* leads to the office equipment making the channels appear busy. Failure of one terminal automatically causes an alarm condition at the distant terminal by removing the carrier supply for a short period.

The carrier fail relay, which is located in the amplifier-regulator mounting, is held operated by a direct current derived from the amplified and rectified carrier pilot frequencies at the output of the flat control unit. Loss of these pilot frequencies releases the relay and causes the alarm circuits in the three terminal mountings to operate. A 3-second delay circuit insures that the "Carrier Fail Alarm" circuits will not operate as the result of very short duration line faults, and also serves as a precaution against accidental restoration due to noise or crosstalk.

Circuits are also provided whereby the system may be transmission tested from either terminal prior to returning the system to service. This feature which is also provided on *O* carrier enables maintenance personnel to check whether trouble has been cleared, before trying to restore the system. It is particularly useful in the case of line trouble which may occur for a short time and then clear itself.

The transmission test is made possible by looping the signaling circuit on the first channel of each group. Signaling tests can then be made from either terminal through the loop, visual indication being given if the test is successful.

A remote restoration feature enables the distant terminal to be restored to service from the near end. Operation of a "System Restore" key restores the near end terminal and transmits signaling tone over the fourth channel of each group. At the distant terminal this tone will cause a thermal relay to operate, restoring that terminal to service. The delay period afforded by this relay prevents faulty restoration due to crosstalk or noise.

Repeater

As indicated in the block diagram, Fig. 17, and in Fig. 18, the repeater is made

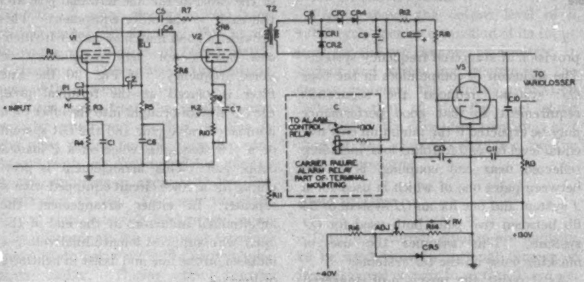

V1
V2
V3
T2
R1
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R16
R18
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
CR1
CR2
CR3
CR4
CR5
L1
P1
K1
K2
+INPUT
TO VARIOLOSSER
C
TO ALARM CONTROL RELAY
130V
CARRIER FAILURE ALARM RELAY PART OF TERMINAL MOUNTING
ADJ
RV
-40V
+130V

Fig. 15. Circuit of flat control unit

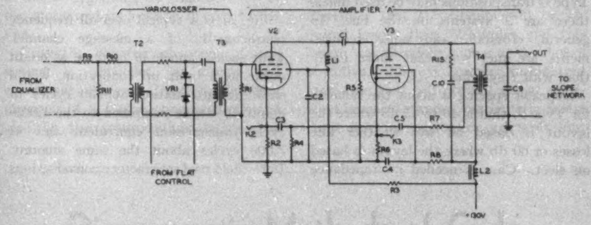

VARIOLOSSER
AMPLIFIER 'A'
FROM EQUALIZER
R9
R10
R11
T2
VR1
T3
R1
V2
V3
C1
L1
C2
R5
R15
C10
T4
C9
OUT
TO SLOPE NETWORK
K2
C3
R2
R4
C5
R7
K3
R6
R8
C4
L2
R3
FROM FLAT CONTROL
+130V

Fig. 16. Circuit of variolosser and amplifier circuit

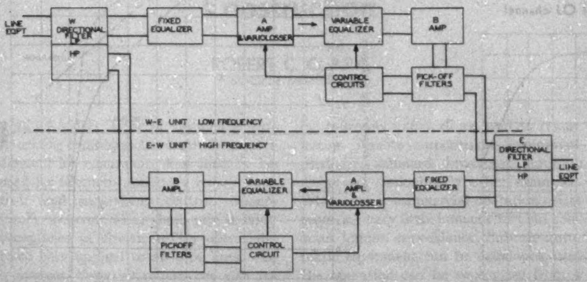

LINE EQPT
W DIRECTIONAL FILTER LP
HP
FIXED EQUALIZER
A AMP & VARIOLOSSER
VARIABLE EQUALIZER
B AMP
CONTROL CIRCUITS
PICK-OFF FILTERS
W-E UNIT LOW FREQUENCY
E-W UNIT HIGH FREQUENCY
E DIRECTIONAL FILTER LP
HP
LINE EQPT
B AMPL
VARIABLE EQUALIZER
A AMPL & VARIOLOSSER
FIXED EQUALIZER
PICKOFF FILTERS
CONTROL CIRCUIT

Fig. 17. OJ repeater

Fig. 18. Type OJ carrier repeater

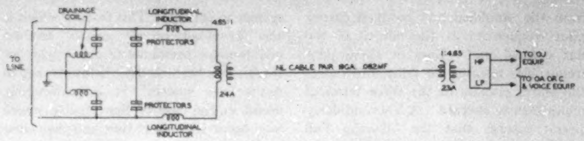


Fig. 19 (above). Line equipment arrangement using a single pair in the entrance cable

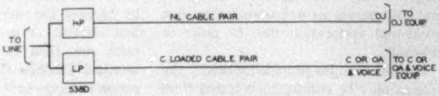


Fig. 20. Line equipment arrangement using one nonloaded and one C-loaded pair in the entrance cable

up of two directional filters and two amplifier-regulator units. The directional filters are alike, requiring simply to be properly oriented in their sockets for the *E* or *W* side of a repeater. The amplifier-regulators are alike except for the pick-off filters and equalizers which are plug-in units designed for the high-frequency (E-W) or low-frequency (W-E) range.

The repeater has a maximum gain of approximately 60 db at the highest frequency. The regulating range is approximately 30-db slope in the high frequency band and 18-db in the low frequency band. Flat regulation at mid-frequency of either band is approximately 45 db.

Lines and Line Equipment

The requirement for crosstalk separation between type *J* systems was established at 59 db. In order to meet this requirement not only was very precise line construction necessary but also the provision of staggered frequency systems. The inclusion of compandors in the type *OJ* systems reduced the separation requirement so that good performance may be expected if the minimum effective equal level far end coupling loss, including reflected near end coupling, is 47 db between pairs one of which is used for a *J* system and one for an *OJ* system or 34 db between two pairs both used for *OJ* systems. This assumes the use of masking noise on the *OJ* systems.

As a result the provison of staggered systems is avoided, simplifying engineering, manufacture and maintenance. Type *O* transpositions may be used unless there are *J* systems on the line. In general, crosstalk and noise requirements are met somewhat more easily than with *J* systems.

Repeater spacing is about the same as for Type *J*, that is 40 to 45 db where the layout is based on wet weather line losses or 60 db where the layout is based on sleet. Care is needed in impedance correction at open-wire to cable junctions, including intermediate cables at river or railroad crossings or through towns, since reflection at such points converts unobjectionable near-end crosstalk into interfering far-end crosstalk.

Of the various line equipment arrangements used at terminal and repeater offices, the two shown in Figs. 19 and 20 are the simplest. In Fig. 19 both the higher frequency *OJ* system and the lower frequency *OA* system, or a *C* system plus voice, are brought in on the same cable pair and separated by a line filter in the office. Autotransformers are used to match the open wire impedance to that of the cable at the line terminal pole and the cable to the office equipment. This provides a good match at carrier frequencies but is not very satisfactory at voice frequency. In Fig. 20 the line filter is placed at the terminal pole, the *OJ* system brought into the office on a nonloaded cable pair and the *OA* system or a *C* system plus voice on a *C* loaded cable pair. This arrangement is preferable for a voice circuit equipped with a repeater. In either arrangement the longitudinal inductors at the end of the open wire suppress longitudinal voltages induced on the line and assist in lightning protection.

System Performance

Fig. 21 is a typical over-all frequency characteristic of a message channel. Four such channels in tandem, as might occur in a built up connection, would show an attenuation at 250 cycles of about 10 db as compared to the 1,000-cycle transmission equivalent and at 3,000 cycles about the same amount. In the field trial satisfactory conversations

Fig. 21 (below). Over-all frequency characteristic of OJ message channel

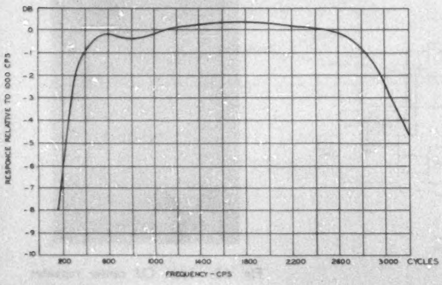
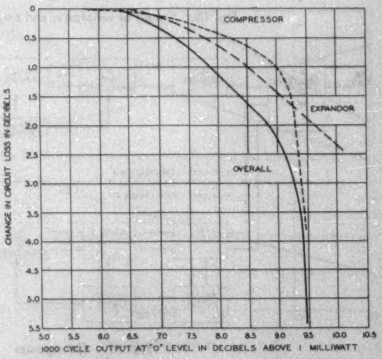


Fig. 22 (right). Limiting load capacity of OJ channel



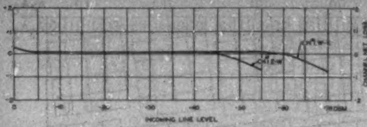


Fig. 23. Transmission regulation of OJ channels with line transmission loss variation

were carried on over eight links in tandem, which is twice the maximum number that might conceivably occur in practice.

The operation of a multichannel carrier system is improved by designing into the channel circuits an amount of power limiting determined by experience. The extent to which power peaks are limited in the *OJ* system is shown in Fig. 22. A very strong talker (zero volume units) will have a peak 10 db above the rms value about 1/10th of 1% of the time. This is about the limiting peak of the system.

Circuit noise from the terminal and repeater equipment are low, so also is line noise during normal operating conditions with no "static" fields due to storm conditions or excessively low line levels due to ice. To mask crosstalk during silent intervals background noise is introduced at the receiving terminal about 57 db below a loud or zero volume units talker. During thunderstorm periods along the open-wire line noise peaks caused by static may rise 15 db or more depending on the severity of the storm and its location relative to the line.

Crosstalk between channels of one system is controlled by circuit and equipment design. Crosstalk between channels of adjacent systems is affected also by the transposition design, line construction and other factors outside the system equipment. Both are greatly reduced by the use of compandors in the *OJ* system. In general the crosstalk from all sources either between channels of one system or between channels of adjacent systems with properly designed and constructed lines should be 60 db below zero level speech.

Fig. 23 shows the output level of a 1,000-cycle tone as a function of its input level for the two channels transmitted at highest and lowest line frequencies, the curves represent the over-all regulation including that of the flat and slope controls in the amplifier regulator and that of the channel regulator in the twin-channel unit.

Field of Use

Open-wire carrier circuits are unlikely to be used between regional centers or from regional centers to sectional centers, since facilities on these back-bone routes will be provided by microwave, coaxial or cable carrier. It seemed improbable, therefore, that *OJ* circuits will be required longer than 500 miles and this length was therefore chosen as the design objective. There is, however, no inherent limitation that will prevent use at somewhat longer distances, particularly where static field intensities are ordinarily low, as in Canada and parts of the United States.

Field Trial

A field trial of the *OJ* carrier system was held during a 6-month period, over Bell Telephone Company of Canada facilities. The trial system, comprising two terminals and two repeaters, was operated over a 150-mile length of *J* transposed line. Other *J*-type carrier systems occupied neighboring positions on the line.

Transmission tests were made over the 150-mile route, and also over longer distances by looping channels together at the 4-wire VF jacks. Regulation tests were carried out by inserting sections of artificial line to simulate the effect of sleet. Crosstalk tests, carried out with the assistance of Bell Telephone Company engineers, indicated that crosstalk into, or from, the *OJ* system was satisfactorily low. In every respect, the field trial results indicated that the design requirements for the *OJ* system had been adequately met.

References

1. TYPE-O CARRIER TELEPHONE, J. A. Coy, E. K. Van Tassel. *AIEE Transactions*, vol. 71, pt. I, 1952 (January 1953 section), pp. 428–37.

2. A TWELVE-CHANNEL CARRIER TELEPHONE SYSTEM FOR OPEN WIRE LINES, B. W. Kendall, H. A. Affel. *Bell System Technical Journal*, New York, N. Y., vol. XVIII, no. 1, Jan. 1939, pp. 119–42.

3. TYPE ON CARRIER TELEPHONE, R. D. Fracassi, H. Kahl. *AIEE Transactions*, vol. 73, pt. I, 1954 (January 1955 section), pp. 713–21.

4. A NEW TELEPHONE CARRIER SYSTEM FOR MEDIUM HAUL CIRCUITS, R. S. Caruthers, H. R. Huntley, W. B. Kahl, Ludwig Pedersen. *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

Diagrammatic Methods of Code Construction

ROBERT C. CLARK
ASSOCIATE MEMBER AIEE

INCREASED USE of automatic operating techniques has created a greater demand for communications links to be used for telemetry, selective remote control, and supervisory control. Supervisory systems which had gained wide acceptance in electric utility operations have been applied to oil pipe lines. In a manner, these operations are also related to ticket reservation equipment used by railroads and airlines, and to transmitter remote supervision equipment used by standard broadcast stations. There are undoubtedly other examples. Whereever a commercial operation exists requiring very little human effort but constant human surveillance, fail-safe automatic equipment can be developed, and the operation can be supervised from a centralized remote-control location.

All remote-control and supervisory circuits impose a common demand on their communications links: the total number of different messages to be transmitted is relatively small, but a transmitted message must be received and decoded at the far end of the link with extremely high reliability. The degree to which this statement is true will depend in each application on the number of different operations or trouble indications involved, and the importance assigned to each of these operations and indications.

Paper 58-997, recommended by the AIEE Telemetering Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-IAS-ARS National Telemetering Conference, Baltimore, Md., June 2–4, 1958. Manuscript submitted May 20, 1957; made available for printing May 20, 1958.

ROBERT C. CLARK is with the General Electric Company, Syracuse, N. Y.

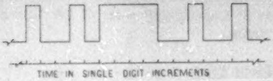


Fig. 1. Typical signaling waveform

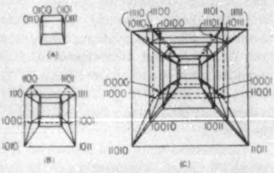


Fig. 2. Three-dimensional coding models

A—Three-digit system
B—Four-digit system
C—Five-digit system

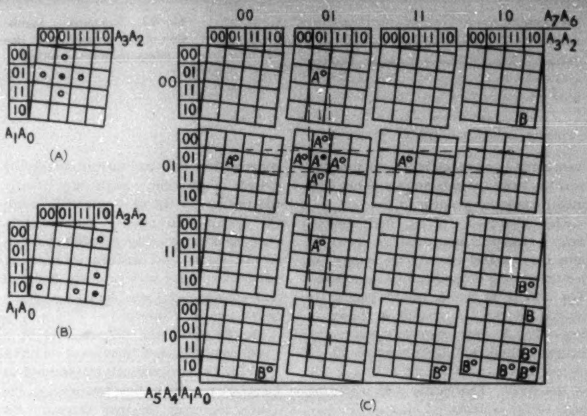


Fig. 3. Examples of single-digit error indication on Karnaugh-Veitch maps. A—Code 0101. B—Code 1010. C—Codes 01010101 (A) and 10101010 (B)

A communications link always has some noise, the presence of which implies a certain probability that the received message is false. For a given average noise power, this error probability can be made arbitrarily small by applying one or more of three techniques:

1. The increase of signal power for more favorable signal-to-noise ratio.
2. The development and use of more efficient schemes of demodulation.
3. Encoding of the messages so that a great amount of noise distortion is required before one message looks like another.

This report deals with the last method of reducing error probability. In particular, it will discuss diagrammatic methods of code construction, and approximate equations whereby the reliability improvement may be evaluated.

The methods discussed here have been evolved as a result of studies relating to time serial binary coding transmitted on a nonreturn-to-zero basis. A typical code transmission is illustrated in Fig. 1. In this coding scheme, a fixed number of equal successive time intervals are assigned "1" or "0" values, independent of adjacent intervals. Each interval, then becomes a digit of a binary number which may be used to represent the message. In general, all messages of a system consist of the same number of time intervals or digits. Teletype systems use this coding, with 5- or 6-digit messages. Actually, these methods of analysis are not limited solely to such codes; they may be used for any coding scheme involving a fixed number of two level inputs, where it is valid to assume that the average correlation between noise of any two separate inputs is zero.

Code-Mapping Techniques

THE 3-DIMENSIONAL MODEL OF N-DIMENSIONAL BINARY CODING SYSTEMS

Hamming[1] introduced the distance protection concept for coding messages to reduce error probability. He has described precisely the procedure to be used for producing single-error detecting or correcting codes and single-error correcting, double-error detecting codes with a minimum addition of code checking and correcting digits to the required message digits. For purposes of explaining the distance protection principle, a 3-dimensional model of a 3-dimensional (or 3-binary-digit) code system, shown in Fig. 2(A), is often used. Each of the three axes of the model is assigned to a single digit. Since each digit can have only two values, all values of the 3-digit function are represented on the eight corners of the cube. Error in a single digit will result in motion along a line of the cube from one corner to the next, e.g., an error in the A_1 digit of code 0010 will result in code 0000. (The order of code digits in this report is identical to that used for conventional number representation, i.e., $A_n \ldots A_2A_1A_0$.)

R. D. Berlin of the General Electric Electric Company has suggested a procedure for extending this model to four or more digits. It will be noted that the corners of the 3-digit model, Fig. 2(A), are labeled with four code digits, the A_3 digit being zero in all cases. This is valid, as the cube can be considered an equipotential surface to any additional code model; then, it is necessary to provide two of different size, assign each cube one of the two values which the fourth digit can assume, and connect corresponding corners of these cubes to form the 4-digit model. This figure may be considered to have 16 corners, representing the 16 possible 4-digit combinations, and four paths intersecting at each corner, representing the four directions by which a single digit of distortion can alter the combination.

The same concept of equipotentiality may be applied to constructing a 5-digit model as shown in Fig. 2(C). There is no limit to the number of digits which can be built into the model, except that the model becomes increasingly cumbersome. This model is used here chiefly to illustrate a technique whereby dimensionally limited systems can be extended to a form usable for studying the multidimensional systems involved in coding.

KARNAUGH-VEITCH MAPPING TECHNIQUES

In their discussions of switching systems and of coding, Keister, et al.,[2] demonstrated that hypercubes and connected hypercubes were a valuable analytic tool. Veitch[3] and Karnaugh[4] have demonstrated a mapping technique based on the hypercube, applying it chiefly to minimization of switching circuits. Caldwell[5] used this map to demonstrate the recognition of symmetric switching functions. O'Brien[6] demonstrated the value of this mapping technique for selecting

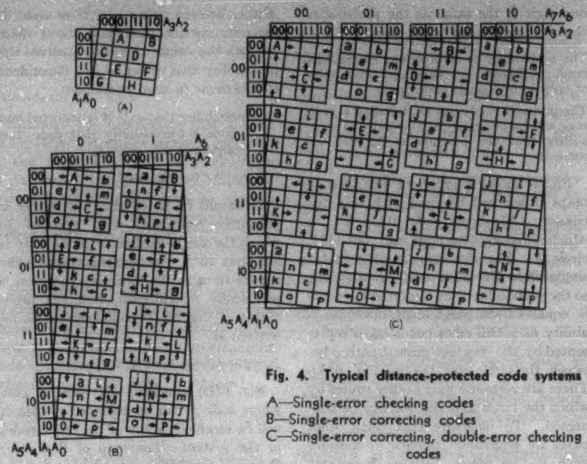


Fig. 4. Typical distance-protected code systems

A—Single-error checking codes
B—Single-error correcting codes
C—Single-error correcting, double-error checking codes

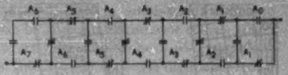


Fig. 5. Symmetric recognition network for code 01010101 and single-digit errors

cyclic coding to be used in a shaft position to digital encoder.

Karnaugh-Veitch maps have characteristics very useful in the analysis of error-protected binary coding. A 2-dimensional map can be drawn representing all combinations of a 4-bit code, with an added valuable feature: If a move is made from one square to an adjacent square in a direction parallel to the map axes, only one digit of the code changes. Fig. 3(A) illustrates this operation. Right-angle motion in any direction away from the solid circle 0101 results in the blank circles 0100, 0111, 0001, and 1101 for changes in the A_0, A_1, A_2, and A_3 digits respectively. Fig. 3(B) illustrates a similar map operation on codes at the corners. Rectilinear motion over the edge of the map is accomplished, assuming that the map is the projected surface of a toroid or a single period of a repetitive chart.

Obviously, a 4-digit map is inadequate to the task of designing a code system. The Karnaugh-Veitch map is not limited in its application to 4-digit codes. Both of the originators[4,5] of the mapping technique have proposed methods of extending the map. A very desirable extension (in the opinion of this writer) is shown in Fig. 3(C). Here the concept of equipotentiality used with the cube discussed in the preceding section is applied to combine 16 of the 4-digit figures into an 8-digit map. In Fig. 3(C), the solid circle labeled *A* represents the code 01010101 while the blank circles labeled *A* represent single-digit errors. The individual maps are canted slightly, so that a straightedge may be used to reduce the problem of identifying squares to be considered as associated between maps; this is illustrated by the dashed lines about code *A* in Fig. 3. By studying the squares connected by the dashed lines, one will note that the rules for single-digit change of code in separated maps are identical to the rules for single-digit change of code within a map. Code *B* illustrates the corner operation of single-digit errors on code 10101010. Again, the extended map operation is identical to the 4-digit map operation.

Given this principle of error recognition, one may now construct a code based on it. If single-error checking is desired, some codes will share the same single-error conditions. Code for single-error checking will be located so that at least two digits change between the codes. This corresponds to $d=2$ in the distance protection concept.[1,2]

A single-error correcting code may be designed if the single-error conditions of that code are not shared with another code. To design such a code system, then, one must locate the codes so that at least three digits change between them. This corresponds to $d=3$ in the distance protection concept.

An additional digit change between codes ($d=4$) will provide a shared second-digit error between codes. The coding will now be single-error correcting, double-error checking.

Fig. 4 will serve to illustrate the distance-protected codes; (A) shows a 4-digit single-error checking code system, and (B) a 7-digit Hamming single-error correcting code system; (C) illustrates an 8-digit Hamming single-error correcting, double-error checking code system. It should be emphasized that the mapping technique is unnecessary for designing the code systems shown in this figure. Clear rules have been stated for their formation.[1,2] (Reference 2 is a readily available book.) Fig. 4 serves as an illustration of the results desired when a distance protected coding system is designed. The mapping technique is most valuable when a system with mixed protection requirements is designed. An example to be given later will show this.

Once a code has been selected using the Karnaugh-Veitch mapping technique, a further advantage may be realized by using its contact minimization features[3,4] to select a switching function. For example, the straight minimization features may be applied to code *A* of Fig. 3 (used for single-error correction) to develop switching function:

$$\begin{aligned}F_A = {} & \bar{A}_7A_6\bar{A}_5A_4\bar{A}_3A_2\bar{A}_1+\bar{A}_7A_6\bar{A}_5A_4\bar{A}_3A_2A_0+\\ & \bar{A}_7A_6\bar{A}_5\bar{A}_3A_2\bar{A}_1A_0+\bar{A}_7A_6A_4\bar{A}_3A_2\bar{A}_1A_0+\\ & \bar{A}_7A_6\bar{A}_5A_4\bar{A}_3\bar{A}_1A_0+\bar{A}_7A_6\bar{A}_5A_4A_2\bar{A}_1A_0+\\ & \bar{A}_7\bar{A}_5A_4\bar{A}_3A_2\bar{A}_1A_0+A_6\bar{A}_5A_4\bar{A}_3A_2\bar{A}_1A_0\end{aligned}$$

Also, the symmetric function technique described by Caldwell[6] may be used to produce the contact configuration of Fig. 5.

Karnaugh-Veitch mapping is still not limited to the 8-digit examples given here. Four of the eight-digit maps may be fitted on $8^1/_2$- by-11 inch sheet of paper for 10-digit code studies. A 17- by 22-inch sheet would permit 12-digit mapping. The map cannot be compressed more than this if space is to be left for necessary notes. A larger sheet of study paper could be provided at this point to extend the map capacity, or 16 of the 12-digit sheets could be combined to provide a 16-digit chapter. The system can be continuously expanded in this fashion; however, the perseverance of most students of the code will undoubtedly be exhuasted at 12 or 16 digits. A 16-digit code map would contain 65,536 code positions and could be used to map a Hamming code with single-error correction and double-error detection for a total of 2,048 codes.

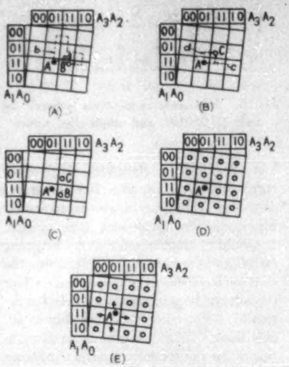


Fig. 6. **Examples of code error situations**

A—Specific single error
B—Specific double error
C—Specific single or double error
D—All possible errors
E—All errors beyond single error correction

ESTIMATING RELIABILITY IMPROVEMENT GAINED BY DISTANCE-PROTECTED CODING

Rice[8,9] has published very complete treatises giving means whereby single-digit error probability of a given detection system can be calculated. There are several sources[10–12] from which the probability that noise will distort a single digit of a code with a particular scheme of sampling can be deduced. Siforov,[13] in his work on codes protected by a distance between them, develops an equation which may be used with the system single-error probability, to determine the code error probability of a given distance protection. His method is applicable to a diagrammatic selection of codes similar to that shown in the preceding section; however, distance in Siforov's work allows the existence of systems in which the distance between codes d is not a whole number. Such a situation may possibly be evaluated in terms of the whole system; however, if the error probability of a single code is to be considered, a fractional distance between is hard to visualize on a coding map. This section will offer a means, compatible to a map method of code synthesis for determining the error probability of a given code.

The means considered here are based largely on a discussion of probability by Goldman.[14] The methods are valid when the code signal is to be distorted by a random disturbance, and when the probability that this distortion occurs in any one digit is the same as the probability that it occurs in any other digit.

The probability of error in any single digit must be known: p_e = average probability of error in any single code digit.

The probability of error in decoding a message will be determined: P_e = average probability that the message is decoded incorrectly.

Fig. 6 shows a group of 4-digit code maps which will be used with this discussion: (A) shows code 0111, dot A, which is distorted by noise into code 1111, circle B. This distortion implies that a single error has been made in digit A_3, so the probability that this error P_e occurs is equal to the single-digit error probability p_e. This error condition is represented by the shortest dashed path a between the circle and the dot Fig. 6(A). There are certainly many other routes by which the circle and the dot can be connected. To illustrate that these other paths are equivalent to path a, consider path b. To follow path b, distortions must occur as follows:

Digit Distorted	From	To
A_1	1	0
A_0	1	0
A_3	0	1
A_0	0	1
A_2	1	0
A_1	0	1
A_2	0	1

By examining these operations, it will be noted that the A_0, A_1, and A_2 digits each undergo two distortions, restoring them to the proper sense, and only the A_3 digit remains in error. Path b is then equivalent to path a. As only one decision is made on each digit, path b is inconsistent with the error which might occur in a single transmission and reception. In summary, the probability of any particular single-error change in a code P_e is equal to the single-digit error probability p_e, and this is indicated by the shortest path ot the adjacent square.

Fig. 6(B) shows a case in which code 0111, dot A, is distorted so that it is received as code 1101 (circle C). Here paths C and D are equally short; however, they are equivalent paths, both indicating that the A_3 and the A_1 digits must be distorted to produce code C. Since the interference was assumed random, and the digit distortion independent, the probability of this code change is evaluated as the probability that two independent events occur (a product) as follows:

$$P_e = p_e X p_e = p_e^2 \text{ for Fig. 6(B)}$$

Fig. 6(C) shows the case in which code A may become either code B or code C. Since in any one trial, only one of these changes can occur, this case involves the probability that either of two dependent events occur (a sum) as follows:

P_e = probability that code A is distorted into code B + probability that code A is distorted into code C

$$= p_e + p_e^2 \text{ for Fig. 6(C)}$$

It should be noted that this paper is primarily concerned with systems in which the single-digit error probability is less than 10^{-3}. Under this condition the square term in the foregoing equation is practically insignificant, and the probability of Fig. 6(C) conditions can be rewritten as

$$P_e = p_e + p_e^2 \approx p_e \text{ for Fig. 6(C)}$$

Fig. 6(D) considers the general case where code A can be distorted so that it will be received as any of the other codes in the system. The map of Fig. 6(D) shows all possible events which can occur in a single transmission and reception of a code, and the sum of all these probabilities is unity. Code A represents the probability that the code is transmitted and received correctly; hence, the probability that the code is distorted into any other position on the map is $P_e = 1$, the probability that the code is received without distortion.

Now, for the code to be received without distortion, each of the four digits must be received correctly. Hence

$$P_e = 1 - (1 - p_e)^4$$

$$P_e < 4p_e \text{ for Fig. 6(D)}$$

Fig. 6(E) shows the situation which must be evaluated when a coding is designed to tolerate single-digit errors, i.e., code A and its adjacent single-error squares are considered to be correct. The probability that the code is distorted into any of the circles represents the code error probability. Two evaluate this, an approximate equation, developed in the Appendix is used:

$$P_e < \frac{N!}{(N-a-1)!\,(a+1)}\, p_e^{a+1}$$

For Fig. 6(E) $N=4$, $a=1$; therefore

$$P_e < \frac{4(3)}{2}\, p_e^2 = 6p_e^2 \text{ for Fig. 6(E)}$$

Given this approximate equation, the code systems in Fig. 4 may be evaluated. The single-error checking system of Fig. 4(A) requires that two probabilities be considered. A code can be received correctly only if it has no distortion; hence the probability that the code is not received correctly, here defined as P_{ne}, will be the probability that at least a single error is made in the 4-digit code ($a=0$, $N=4$):

$P_{ne}<4p_e$ for Fig. 4(A).

Although the probability that a code in Fig. 4(A) will not be received correctly is the probability that a single digit is in error, the probability that the code will be distorted into an erroneous code requires the consideration of 2-digit errors ($N=4$, $a=1$):

$P_e<6p_e^2$ for Fig. 4(A)

For the 7-digit single-error correcting system of Fig. 4(B), the probability that the code is not received correctly is the same as the probability that an erroneous code is received. This requires a 2-digit error in the 7-digit code ($N=7$, $a=1$):

$P_e<\frac{7(6)}{2}p_e^2=21p_e^2$ for Fig. 4(B)

For a single-error checking, double-error checking code system such as that shown in Fig. 4(C), two probabilities must again be considered. First, there is the probability that the code will not be received correctly, due to 2-digit errors in the 8-digit code ($N=8$, $a=1$):

$P_{ne}<\frac{8(7)}{2}p_e^2=28p_e^2$ for Fig. 4(C)

Next there is the probability that the code will be received erroneously, due to an error of at least three digits in the 8-digit code:

$P_e<\frac{8(7)(6)}{3!}p_e^3=56p_e^3$ for Fig. 4(C)

Thus far in this consideration of code error probability, no considertion has been taken of the fact that, due to the presence of noise, the receiving decoder is not in synchronism with the received code. This consideration can be extremely important in a serial bindary coding. For example, if the means of synchronism is the addition of a single-start digit at the the beginning of a code transmission, the probability that the message is received in error (defined here as P_{em}) will be determined by two dependent error probabilities: the probability that the code digits are in error (P_e), and the probability that the synchronizing message is in error (here defined as P_{es}):

$P_{em}=P_e+P_{es}$

for codes dependent on a synchronizing message. Again, for the single-digit synchronizing message, the error probability is equal to the single-digit error probability:

$P_{em}=P_e+p_e$

for codes with a single-digit synchronizing message.

The consideration of synchronizing error probability is introduced into this section largely so that two related conclusions concerning reliable coding systems can be drawn as follows:

1. The reliability which can be secured by distance protected coding is often limited by means of synchronism.

2. If a code system is to be designed to tolerate a code error of K digits, the synchronizing message should also be designed to tolerate a synchronizing error of K digits.

An Application of Diagrammatic Techniques to Code Design

Use of the Karnaugh-Veitch map is designing a protected code system does not guarantee that an optimum coding will be produced, but it does provide considerable help in assembling mixed code systems. An example will illustrate this.

Assume that a communications link is available which can be used for code transmission with an error probability of 10^{-4}. Assume also that synchronizing means are available in the detection equipment which will provide synchronous detection with timing error probability of 10^{-7}. For this link and detector, it is desired that telemeter codes indicating discrete 10% levels from 0.0% to 100% be constructed with a probability of less than 10^{-6}, and that the received level is more than 10% of full scale different from the transmitted signal. A minimum of ten additional signals is desired, which will be used to transmit supervisory functions with an error probability of less than 10^{-6}.

To evaluate the range of protection required, substitute the specifications into the approximate equation used for Fig. 6(E) in the preceding section.

$$P_e=\frac{N!}{(N-a-1)!\,(a+1)!}p_e^{a+1}$$

and p_e 10^{-6} by specification; therefore

$$10^{-6}\geq\frac{N!}{(N-a-1)!\,(a+1)!}(10^{-4})^{a+1}$$

For $a=1$, any N less than 14 satisfies the inequality. Fourteen digits, providing a coding space for 16,384 unprotected codes, should be more than is required. Hence the system will be designed with $a=1$.

The specification for telemetering codes permits 10% errors, with no specified probability, but does require that errors larger than 10% be held to a probability of 10^{-6}. Such specification implies that the codes are distance-protected, but that codes adjacent in amplitude value may also be adjacent in the code space. Such an arrangement can be made, where one of the N possible single-digit errors of a given code may be construed as the next higher code, and another of the N possible single-digit errors may be construed as the next lower code. This construction will involve a region for each code consisting of one element of space for the true code and N-2 elements for the single-digit errors of a total of N-1 elements. As, an exception, the first and last codes can err in only one direction, and thus must have N-1 elements for single-digit errors, or a total of N elements. For maximized coding in this scheme, an N-digit code space may contain a total of $(2^N\text{-}2)/(N\text{-}1)$ such codes. A 6-digit space will provide a maximum of 12 such codes, whereas only 11 are required by the system. Fig. 7(A) shows a 6-digit map containing 11 such codes, A through K. Capital letters are used to indicate the actual coding positions. Arrows extending from these positions are used to indicate automatically correctible single-digit errors occupying the same 4-digit map as the coding position. Lower-case letters are used to indicate automatically correctible single-digit errors not on the same 4-digit map as the coding position.

This 6-digit telemetering code can be made part of an 8-digit system including the supervisory codes. The influence of the 6-digit telemetering codes on the 8-digit map is indicated in Fig. 7(B). The spaces on this expanded map, which will be occupied by the telemetering codes of Fig. 7(A), are crossed out, and additional single-error versions of the telemetering codes produced in the expanded map are indicated by lower-case letters. To preserve the required protection, it is also essential that added codes be spaced more than two digits away from the telemetering codes. Two-error versions of the telemetering codes are indicated by circles so that these spaces will not be selected as actual code locations in the expanded system.

In the remaining coding spaces, supervisory codes are selected to operate on the following principle: The desired supervisory function will be transmitted, and repeated at intervals until the distant station returns a signal verifying that it has received permissible supervisory code. The verification code ZZ is first selected in a location to be four error digits away from all other code positions, and is designed to be single-error correcting, double-error checking. The 2-digit error circles for ZZ are underlined to indicate that no code position may be located closer than two-digit errors from them. Actual supervisory codes are designed to be single-error checking only, except that where code locations permit, the 2-digit error spaces determined by the telemetering codes are used as single-error correcting spaces for the supervising codes. These supervisory codes (L to Z and ZZ) are mapped on Fig. 7(C), and combined with the telemetering codes in the complete system map on Fig. 7(D).

There are several error probabilities to be considered in the code system shown in Fig. 7(D). Telemetering codes A and K include one single-error possibility that will cause 10% error; hence the probability that they are 10% in error is 10^{-4}. Codes B through J include two such single error possibilities, and thus have a 10% error probability of 2×10^{-4}. For all the telemetering codes the probability of more than 10% is the probability that more than one digit of the 8-digit code is distorted ($N=8$, $a=1$):

$$P_e < \frac{8(7)}{2}(10^{-4})^2$$

$$P_e < 3\times10^{-7}$$

for error of 10% in codes A through K.

The supervisory codes cannot be corrected with errors in any one of six or eight of their digits; hence the probability that the code is not received is 6×10^{-4} or 8×10^{-4}. Due to the selected operation, this simply implies that these codes will have to be repeated six or eight times

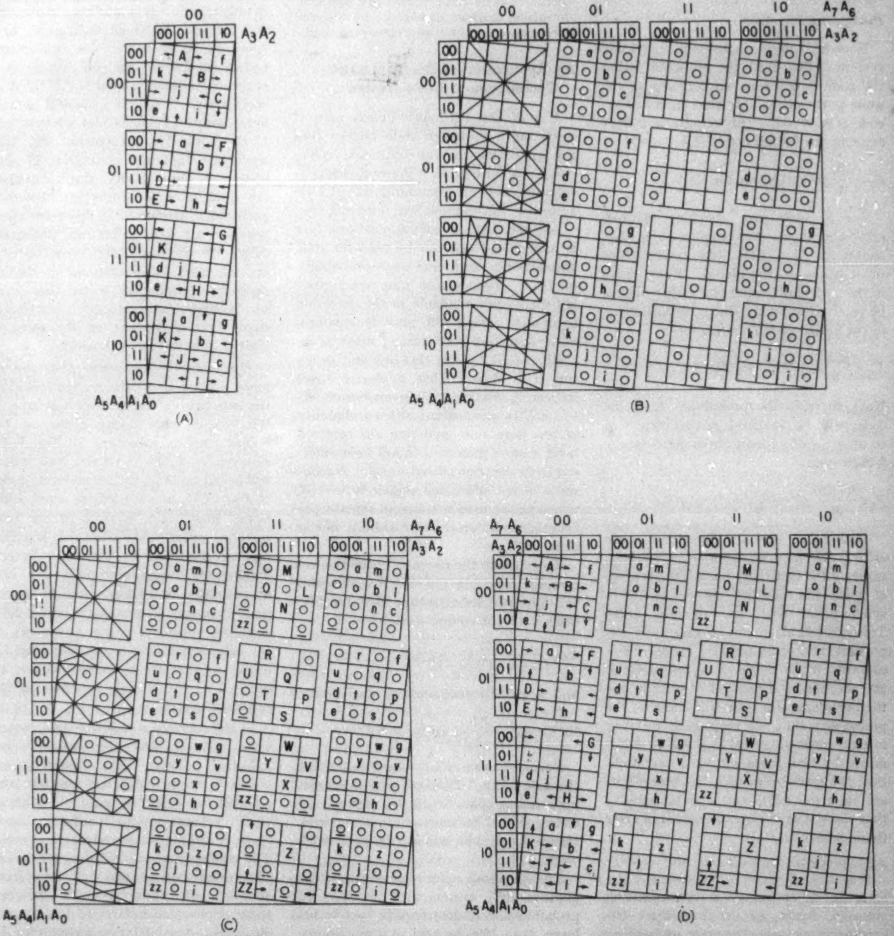


Fig. 7. Construction of a protected code system. A—Basic telemetering code. B—Effect of telemetering code on 8-digit code space. C—Supervisory code added to 8-digit code space. D—Combined telemetering and supervisory code system

per each 10,000 such transmissions. The more important probability, that a supervisory code is distorted into another valid code, requires more than one error:

$$P_e < \frac{8(7)}{2}(10^{-4})^2$$

$$P_e < 3\times 10^{-7}$$

for one supervisory code distorted into another valid code.

The verification code is single-error correcting; hence the probability that it is not received is 3×10^{-7}, as demonstrated for previous considerations permitting single errors. Due to its double-error checking feature, an error of more than two digits is required before the verification code becomes a different code, or vice versa:

$$P_e < \frac{8(7)6}{3!}(10^{-4})^3$$

$$P_e < 6\times 10^{-11}$$

for erroneous decoding in the verification codes.

The coding offered in Fig. 7(A) does not use the code space in the most efficient possible manner. Hamming offers an equation indicating that a maximum of 28 single-error correcting codes can be located in an eight-digit code space. Fig. 7(D) is, with the exception of the verification code, less than single-error correcting and does contain only 27 codes. On the other hand, the largest reported single-error construction of an 8-digit code, to the writer's knowledge,[13] contained only 19 messages. An 8-digit code space contains 256 code locations, yielding a total of 2^{256} or approximately 10^{77} functional combinations. To fit sets of these combinations perfectly is a complex problem. The diagrammatic technique shown in this report facilitates the assembly of code systems involving special protection considerations; further direction is needed to make the assembled systems as efficient as possible.

Appendix. Development of the Approximate Formula for Evaluating Error Possibility in Distance-Protected Code

When a code of an N-digit code system is transmitted, there is a finite probability that the same code will be received, or that any of the other 2^N-1 codes in the system will be received. The sum of the probability of all these possible events is unity; this may be expressed using the Bernoulli distribution law.[13]

$$1 = \sum_{R=0}^{N} C_R{}^N P_e{}^R (1-p_e)^{N-R} \qquad (1)$$

where $C_R{}^N$ is the combination of N things taken R at a time.

Note that each term of the foregoing summation states the probability that the received code will be in the group of codes located at a radius R, from the intended code (i.e., the probability of exactly R digit errors in the communication). When a coding is designed to tolerate an error of A digits, the summation in equation 1 can be divided two separate summations:

$$1 = \left[\sum_{R=0}^{a} C_R{}^N p_e{}^R (1-p_e)^{N-R}\right] + \left[\sum_{R=a+1}^{N} C_R{}^N p_e{}^R (1-p_e)^{N-R}\right] \qquad (1A)$$

The first summation represents the probability that the received code is within the specified tolerance; the second the probability that it is outside the specified tolerance. Either of these summations may be used to calculate the code error probability P_e:

$$p_e = 1 - \sum_{R=0}^{a} C_R{}^N p_e{}^R (1-p_e)^{N-R} \qquad (2)$$

or

$$P_e = \sum_{R=a+1}^{N} C_R{}^N p_e{}^R (1-p_e)^{N-R} \qquad (3)$$

Equation 3 may be further expanded:

$$P_e = \sum_{R=a+1}^{N} C_R{}^N p_e{}^R \left[\sum_{j=0}^{N-R} C_j{}^{N-R} (-p_e)^j\right] \qquad (4)$$

Equation 4 may be expanded and recollected in corresponding powers of p_e to yield

$$P_e = C_{a+1}{}^N p_e{}^{a+1} \left[\sum_{K=0}^{N-(a+1)} \frac{a+1}{a+1+K}\, C_K{}^{N-(a+1)} (-p_e)^K\right] \qquad (5)$$

The summation in equation 5 is an alternating series. The first term of the series is unity. Terms in the series will diminish from the mth term on, provided that mth term can be divided by the term immediately preceding it to yield a quotient less than unity, i.e.,

$$\frac{(a+m)(N-a+m)}{(a+m+1)m}\, p_e < 1 \qquad (6)$$

Now, the object of this investigation is the development of a simple formula approximating the exact relation given in equation 5. Such an approximation will be based on the following concepts:

1. The simplest approximation of equation 5 would involve only the first term of the summation.

2. Since error proability is stated to one significant figure, the first term of the summation will determine the error probability to sufficient accuracy provided the second term is less than 5% of the first term, i.e.,

$$\frac{(a+1)(N-a+1)}{(a+2)}\, p_e \le 0.05 \qquad (7)$$

If the conditions of equation 7 are fulfilled

$$P_e < C_{a+1}{}^N p_e{}^{a+1} \qquad (8)$$

$$P_e \approx C_{a+1}{}^N p_e{}^{a+1} \qquad (9)$$

i.e., the approximate formula given in equation 9 provides a means of calculating error probability which will be pessimistic, but less than 5% pessimistic. The following table illustrates the restrictions imposed on the use of the approximate formula by the conditions in equation 7:

Error Tolerance (a)	Maximum Code Length (N)		
	$p_e=0.01$	$p_e=0.001$	$p_e=0.0001$
0	9	99	999
1	7	75	750
2	7	67	667
3	8	63	627
4	9	63	603
5	9	62	587

References

1. Error Detecting and Error Correction Codes, R. W. Hamming. *Bell System Technical Journal*, New York, N. Y., vol. 26, no. 2, Apr. 1950, pp. 147–60.

2. The Design of Switching Circuits (book), W. Keister, A. E. Ritchie, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1951. (Particularly chap. 8, sect. 8.7, and chap. 12, sect. 12.4.)

3. A Chart Method for Simplifying Truth Functions, E. W. Veitch. *Proceedings*, Association for Computing Machinery, Pittsburgh, Pa., May 2–3, 1952.

4. The Map Method for Synthesis of Combinational Logic Circuits, M. Karnaugh. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 593–99.

5. The Recognition and Identification of Symmetric Switching Functions, S. H. Caldwell. *Ibid.*, vol. 73, pt. I, May 1954, pp. 142–47.

6. Cyclic Decimal Codes for Analogue to Digital Converters, Joseph A. O'Brien. *Ibid.*, vol. 75, pt. I, May 1956, pp. 120–22.

8. Mathematical Analysis of Random Noise, S. O. Rice. *Bell System Technical Journal*, vols. 23, 24. (Particularly Part I, July 1944, and Part III, Jan. 1945.)

9. A First Look at Random Noise, S. O. Rice. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 128–31.

10. The Integration Method of Detection and Its Application to the Transmission of Teletype Signals, H. F. Harmuth. *Report AD-69006*, Armed Services Technical Agency, Washington, D. C. 1955 (not classified). (Section II-2 and associated figures contain information on error probability for several schemes of amplitude sampling and integration detection; Sections III-4 and III-5 contain special relations to be used for evaluating error probability when distance protected codes are used with the author's method of integration detection.)

11. Phase Shift Radio Teletype, J. P. Costas. *Report R56ELS28*, General Electric Technical Information Series, Schenectady, N. Y., 1956.

12. A Predicted Wave Radio Teletype System, M. L. Doelz, E. T. Heald. *Convention Record*, Institute of Radio Engineers, New York, N. Y., 1954, pt. 8.

13. On Noise Stability of a System with Correcting Codes, V. I. Siforov. (English translation by M. D. Friedman.) *Supplementary Transactions*, Symposium on Information Theory, Massachusetts Institute of Technology, Cambridge, Mass., 1956.

14. Frequency Analysis Modulation and Noise (book), S. Goldman. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, particularly pp. 286–93.

SAGE Data System Considerations

R. G. ENTICKNAP
NONMEMBER AIEE

E. F. SCHUSTER
NONMEMBER AIEE

A SYSTEM for the transmission of binary digital data over telephone circuits has been developed for the SAGE (semiautomatic ground environment) System of air defense. This data transmission system had to meet stringent and varied requirements for data rate, error rate, general applicability, cost, and not least importantly, reliability in the face of enemy action. The nature of these and other requirements are discussed together with some of the history of the development program which led to the present solution. A detailed description of the data-transmission system now being installed throughout the nation is given. The conclusion contains a discussion of problems still remaining and some opinions of future trends in data-transmission systems for air defense and other applications.

The Need for a Data Transmission System

In the year 1951, the "Charles Committee" was appointed by the Joint Chiefs of Staff to study the problem of air defense of the United States and to make recommendations concerning the nature of possible solutions and the means of achieving the desired goal.[1]

The future system of air defense that was proposed and subsequently accepted by the Air Force, involved the concept of a small number of centers where large-scale data processing and display machinery would be located and where instructions and control information would be generated.

Radars and other information sources would feed information into these centers now called direction centers (DCs) and the DCs would send to interceptors and other weapons the messages involved in their deployment and control.

The communication requirements for the transmission of this information were revolutionary. SAGE was planned to use the existing radars and weapons, together with many new ones, all connected to the new DC. The location of these radars was fixed by the requirements for complete air surveillance coverage and by the fundamental limitations in the performance of line-of-sight radars, and these were to be augmented to insure adequate low-altitude coverage. The locations of the DCs were largely determined by logistic factors and the paramount need for avoiding areas that might be considered prime targets for enemy attack.

Existing interceptor bases and new missile-launching bases that were to be integrated into the SAGE System were, of necessity, located on a geographic basis, to insure that an optimum distribution of weapons to protect prime targets, or expected attack routes, would be maintained.

Radio coverage for primary weapon control instructions had to match that of the radars if the weapons were not to be restricted in their operational capability. Consequently a multiplicity of such sites had to be provided, the locations being determined by the fundamental limitations of the radio equipment.

Since data were to be processed at the DC the output of the radars had necessarily to be conveyed to the DC, and with a minimum of delay. Similarly large quantities of data had to be conveyed from the DC to weapons control radio sites, and to adjacent data-processing centers to insure adequate handling of air battles which traversed the boundaries of the center control areas.

This need for transmission of automatically processed data, over long distances, between machines, in contrast to voice communications between human beings, required fundamental changes in air defense communications engineering.

Particularly applicable to the radar case, one technical solution would have been to use wide-band microwave circuits to transmit the radar video output to the DC. But this solution, apart from its prohibitive costs, would have been hopelessly inefficient because the data-processing machinery deals only with information which describes discrete radar returns. This discrete information can on the average be contained in a band-width which is approximately 1/1,000th of the video output of the radar receiver.

The more desirable system therefore involves machinery at the radar to quantize and store the processed radar data and a communication system of capacity sufficient to pass these data at no less than the average rate at which it is generated.

Outputs of the data-processing DC are originated in the form of discrete digital data. These outputs require a transmission system capable of handling at least the average rate of generation, preceded by machinery with storage to handle the transition from generation rate to average transmission rate and capable of encoding information in a form suitable for transmission.

Upon detailed examination of these and other like communication requirements, the need for a new data-transmission system became apparent. Existing data systems such as teletypewriter, were of too low a capacity and when used in multiple, too costly, and wide-band microwave video systems were unnecessarily high in capacity and very costly to install.

Historical Development

Following the recommendations of the Charles Committee, the Massachusetts Institute of Technology (MIT) formed a new laboratory to carry on research and development in the field of air defense. One of the first tasks of the Lincoln Laboratory, as it was named, was to demonstrate the principles of the proposed new semiautomatic air defense system by means of an experimental model.

Known today as the Cape Cod System, the model was set up in Eastern Massachusetts, using the Whirlwind computer in MIT's Barta Building as the basis of an air defense direction center. A number of small radars were installed at 20 to 50 miles distance from the center. Interceptor and target aircraft, based at Hanscom Air Force Base, were provided by the Air Force for the experiment.

The early data-processing machinery at these radars and the data-transmission system stemmed from work originated at the Air Force Cambridge Research

Vol.
Pt.
195

Paper 58-302, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted November 21, 1957; made available for printing September 10, 1958.

R. G. ENTICKNAP is with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., and E. F. SCHUSTER is with the Western Electric Company, Inc., ADES, New York, N. Y.

The authors wish to acknowledge the many contributions to this field, not only by members of their respective organizations, but also by individuals in the United States Air Force associated with the SAGE System. Contributions from industry were numerous and came principally from the American Telephone and Telegraph Company, The Bell Telephone Laboratories, Inc., and the International Business Machines Corporation.

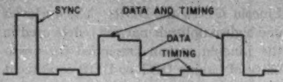


Fig. 1 (left). Modulation waveform

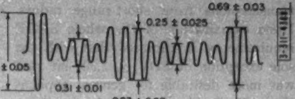


Fig. 2 (right). Idealized phone line signal

Center, at that time also located in Cambridge, Mass.

The first radar data-processing device used was called slowed down video (SDV) and it quantized the radar picture into range circles, and into azimuth sectors. The data-transmission system used a leased conventional voice telephone circuit as a transmission medium, over which was sent amplitude-modulated signals; see Figs. 1 and 2.

Since this processing and transmission was carried out in real time, the data rate was a function of the radar scan rate, and for these radars 1,600 bits/second was adequate. The inevitable small variations in radar scan rate were taken care of, not by varying the transmission rate, but by allowing variations in the number of range boxes in each discrete azimuth sector scan.

All the telephone circuits in the original Cape Cod System were made up of voice frequency facilities, and although certain of these proved too noisy and had to be changed, the data-processing and transmission systems proved adequate for the purposes of the early experiments with short-range radars.

When it was decided to enlarge the Cape Cod System, particularly to include longer range radars, a number of difficulties with the data-processing and transmission system became apparent.

To fulfill the need for an improved data-processing and -transmission system for long-range radars, a machine known as fine grain data (FGD) was developed. This device, a small computer in itself, works on entirely different principles from the SDV. In FGD, a radar echo is treated in its several separate component pulse returns and a target position determined as the weighted average of all these separate pulses. The *x* and *y* co-ordinates of this target position are determined to an accuracy much greater than with SDV and together with the time of occurrence are stored in binary form in a storage register. Two 1,300-bit circuits are used to transmit this data to the DC. In addition to handling search radar returns, identity IFF (Identification Friend or Foe) radar and height information from height radars are also stored by the FGD machine. This information is transmitted over the same two data circuits in time multiplexed form. The FGD equipment in production for SAGE use is called the Coordinate Data Transmitting Set.

In addition to search radars, other components of the Cape Cod System had need of a data-transmission system. At first height finding was carried out manually, requests and data being passed by voice to and from the center and the height-finder radars. This height system had to be improved if it was to be compatible with the performance of the rest of the system. An experimental ground-to-air data link (G/A DL) was used in the Cape Cod System for passing computer generated instructions to interceptor aircraft. The transmitter was remote from the computer and a data-transmission system was needed to convey the instructions.

The early Cape Cod experiment completely fulfilled its purpose however. It successfully demonstrated the technical feasibility and operational advantages of this revolutionary concept of an air defense system. It was instrumental in guiding the Air Force decision to proceed with a nation-wide installation. This decision demanded that the Lincoln Laboratory apply its major effort to the application of the experimental designs and equipment to a large-scale permanent installation. Not least of the many problems was that of designing and planning a new data-transmission system, to be installed on a vast scale, and with problems hitherto unknown in a nation-wide communications system.

Technical Requirements

In the consideration of requirements for the data-transmission system for SAGE, the following appeared paramount.

1. There should ideally be one "standard" system used throughout the SAGE System for all applications.
2. It should use as a transmission medium a type of circuit readily available throughout the country.
3. The incidence of errors introduced by the data-transmission system should be a minimum consistent with item 2.
4. The capacity for transmitting data (data rate) should be as high as possible consistent with items 2 and 3.

STANDARDIZATION

The general advantages here are obvious and almost need no mention. In the SAGE System there is an outstanding technical, rather than the usual logistical, advantage of standardization. A common standard of data rate and type of signal makes possible the sharing of computer input or output sections which cannot be fully utilized by one particular service.

TRANSMISSION MEDIA

It has been previously mentioned that in the early Cape Cod System conventional leased voice circuits were used for data transmission. It had been hoped that such circuits could be used throughout, but upon investigation, voice circuits were found to be quite unsuitable to meet SAGE data-transmission requirements.

First, there is no specification available to the user, of the characteristics of a voice circuit, other than it be capable, when suitably terminated, of satisfactorily transmitting intelligible speech. The data rate would be limited by the worst case of amplitude and phase distortion characteristics likely to be encountered, and furthermore, the worst case would have to be found by measurement. Second, with voice transmission, the noise on the circuit, particularly, impulse noise, is far less important than with data transmission, and there were no data available on the impulse noise characteristics, or their distribution of conventional voice circuits. Third, it is in the nature of telephone company operations that leased circuits may be changed in their make-up, or rerouted at any time to meet prevailing conditions, so that even if a voice circuit was established and found to be satisfactory for data transmission, it would not necessarily remain so indefinitely.

It became necessary then to establish a new class of circuit. Because of the need for general availability without extensive construction, the basis of the circuit had to be the same facilities used for voice transmission.

DATA RATE

It was stated earlier that 1,600 bits/second was the rate chosen for SDV transmissions from short-range radars. Upon examination of the whole SAGE requirements, it appeared that with the sole exception of SDV, a 1,300-bit rate was more desirable for general application. Used either as a single circuit, or in multiples of no more than two, all SAGE requirements could be met, including the FGD system developed for the processing of data at long-range radars. The lower bit rate was more compatible with voice circuit bandwidths and was less demanding in envelope delay equalization; also, it promised some small improvement in error rate.

ERROR RATE

In the Cape Cod System, it was observed that the noise on telephone circuits was generally of two types: bandwidth limited white noise and impulse noise. Of the two, the impulse noise contributed the bulk of the errors, since the normal signal level (0 decibels referred to milliwatts to the sync or start pulse) was well above the mean noise level. Impulse noise spikes were often of similar or greater magnitude than the level of data pulses.

A noise impulse, after being shaped by the circuit, often appeared as a false bit of data or might cancel a real bit of data, depending upon its time of occurrence and its phase relationship. If the error rate was to be minimized, then attention would need to be focused on reduction of impulse noise, rather than on improvement of signal/white noise ratio.

In early 1954 a committee was formed to specify the characteristics required for a data circuit to meet the requirements in the preceding.

In Appendix I the specifications are given that the committee decided upon for a data circuit, which appeared to satisfy all requirements. The characteristics of the signal to be sent over this circuit were also defined and are given in Appendix II and Figs. 1 and 2.

A number of data circuits were provided to replace the voice circuits in use in the Cape Cod System and the anticipated improvement in performance over the conventional voice circuits was immediately apparent.

The terminal devices, modulators, and demodulators that were used in the Lincoln data-transmission system were called digital data transmitters (DDTs) and digital data receivers (DDRs) respectively. The input to the DDT consisted of unidirectional pulses for sync and data signals and of a continuous sine wave for timing. The output of the DDR consisted of unidirectional pulses for all three signals, sync, data, and timing. The amplitude and duration of these input and output pulses were chosen to make them directly applicable to the digital logic circuits of the data-processing machinery.

The "on line" signal transmitted by the DDT Fig. 2, Appendix II, utilized a basic carrier onto which the signal waveform was modulated.

This particular design of DDTs and DDRs evolved directly from experience with the specific telephone circuits used in the Cape Cod System and with loop circuits in the laboratory for test purposes.

With increasing experience gained with a larger variety of circuits, came the further realization that it was impractical to consider the DDT, DDR, and telephone circuit as separate entities, but because of the many types of telephone circuits and continued improvement in telephone facilities, they should be considered as a complete system providing digital data transmission between two points.

Although Lincoln designs of DDTs and DDRs appeared to be adequate, it was recognized that separate ownership of the terminal equipment would inhibit changes which could utilize important inventions in the field of telephone communications. In addition, requirements had been stated by the Air Force for rapid restoration of data circuits in the event of failure. The optimum solution required that trouble sensing devices and automatic switching be included as part of the terminal DDR equipment. Further, from the point of view of maintenance, it was considered essential that test signals and devices be included which would check the data circuits as well as the DDTs and DDRs.

Accordingly, agreements were reached on the characteristics of a complete data service which would include both terminal equipment and transmission media.[2,3]

A standard signal specification was agreed on, such that the synchronization, data, and timing signals are all delivered to a Data Service as a standard sine wave or dipulses and the new DDT transmits these signals in a manner optimum for transmission over the particular facility which is being employed. At the receiving end, the new DDR converts the complex on-line signal into three standard sine wave or dipulse signals for delivery of sync, data, and timing information to the user.

The specifications for input signals to DDTs and output signals from DDRs are given in Figs. 3 and 4. These differ considerably from those used with the Lincoln DDTs and DDRs. The d-c levels and high-amplitude narrow pulses used in the Lincoln system precluded the possibility of having DDT or DDR equipment remotely located from the data sources or data receiving equipment except when interconnected by coaxial cable. In addition, equalization might have been required. Further, the necessary switching of circuits for restoration, or patching to other input or output equipment was complicated enormously with the Lincoln-type wave forms.

In order to obviate these difficulties, input and output signal wave forms have been established for the Digital Data Service, which allow transmission via regular telephone facilities and house cables, and which allow switching and patching via conventional equipment. These input and output signals have been adopted as standard for the SAGE System.

The specified error rate for the Data Service is identical to that specified in Appendix I for data circuit, i.e., not more than one impulse noise spike, (or bit error), per minute. This, at a 1,300 bit/second data rate, is approximately 1 in 10^5 bits. It has been long recognized that this particular specification is unsatisfactory because it does not specify distribution either on the circuit in question or in relation to other data circuits. Telephone circuit noise is not generally a purely random phenomenon but often occurs in bursts, in which a large number of errors is concentrated in short time intervals; the effect of this noise on system operations, is more dependent on its time distribution than on its frequency of occurrence. It is not, therefore, sufficient to specify only a random rate of permissible errors; the length of the intervals during which these normal rates may be exceeded and the minimum time between these intervals must be considered.

The effect of errors on the SAGE System can only be considered for a whole class of like inputs or outputs, and not on an individual circuit basis. It has been calculated that for all the gap filler radar inputs to an average center, a random error rate of one per 10^5 bits would result in a 1% loss of automatic initial pick-up capacity. This is well within the system capability but if the total errors from gap filler circuits were distributed such that one in 10^4 bits were in error for a given period, then some 10% of the initial automatic pick-up capacity of the computer might be wasted during that time. Similar calculations have been made for other classes of inputs and outputs which

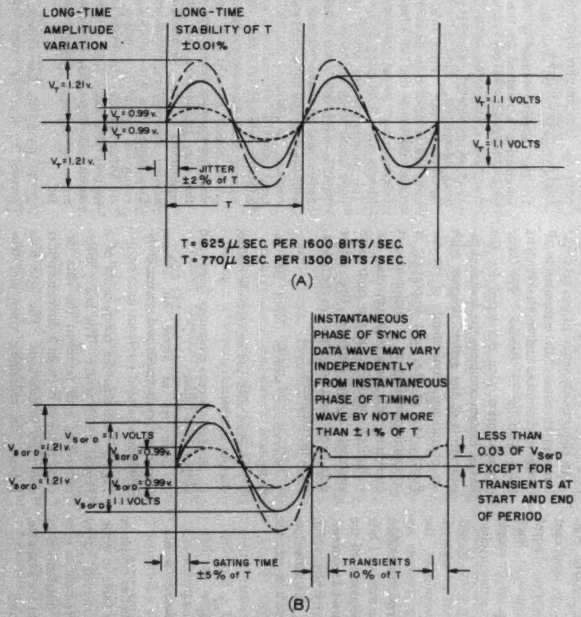

LONG-TIME AMPLITUDE VARIATION
LONG-TIME STABILITY OF T ±0.01%
V_T = 1.21 v.
V_T = 0.99 v.
V_T = 0.99 v.
V_T = 1.21 v.
V_T = 1.1 VOLTS
V_T = 1.1 VOLTS
JITTER ±2% of T
T
T = 625 μ SEC. PER 1600 BITS/SEC.
T = 770 μ SEC. PER 1300 BITS/SEC.
(A)
INSTANTANEOUS PHASE OF SYNC OR DATA WAVE MAY VARY INDEPENDENTLY FROM INSTANTANEOUS PHASE OF TIMING WAVE BY NOT MORE THAN ± 1% OF T
$V_{S\ or\ D}$ = 1.21 v.
$V_{S\ or\ D}$ = 1.1 VOLTS
$V_{S\ or\ D}$ = 0.99 v.
$V_{S\ or\ D}$ = 0.99 v.
$V_{S\ or\ D}$ = 1.1 VOLTS
$V_{S\ or\ D}$ = 1.21 v.
LESS THAN 0.03 OF $V_{S\ or\ D}$ EXCEPT FOR TRANSIENTS AT START AND END OF PERIOD
GATING TIME ±5% of T
TRANSIENTS 10% of T
(B)

Fig. 3. SAGE signals to be delivered from data source to interconnection unit

A—Timing signal
B—Sync or data signal
Mark Space
Nominal loop impedances = 600 ohms, balanced

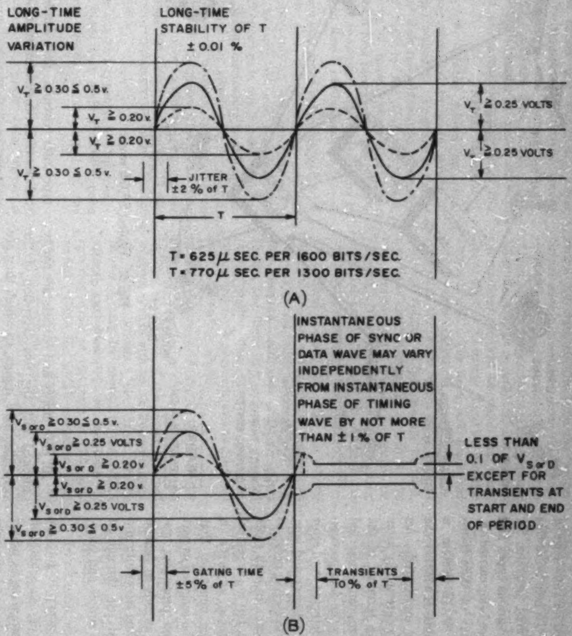

LONG-TIME AMPLITUDE VARIATION
LONG-TIME STABILITY OF T ± 0.01 %
$V_T \geq 0.30 \leq 0.5$ v.
$V_T \geq 0.20$ v.
$V_T \geq 0.20$ v.
$V_T \geq 0.30 \leq 0.5$ v.
$V_T \geq 0.25$ VOLTS
$V_T \geq 0.25$ VOLTS
JITTER ±2 % of T
T
T = 625 μ SEC. PER 1600 BITS/SEC.
T = 770 μ SEC. PER 1300 BITS/SEC.
(A)
INSTANTANEOUS PHASE OF SYNC OR DATA WAVE MAY VARY INDEPENDENTLY FROM INSTANTANEOUS PHASE OF TIMING WAVE BY NOT MORE THAN ± 1% OF T
$V_{S\ or\ D} \geq 0.30 \leq 0.5$ v.
$V_{S\ or\ D} \geq 0.25$ VOLTS
$V_{S\ or\ D} \geq 0.20$ v
$V_{S\ or\ D} \geq 0.20$ v.
$V_{S\ or\ D} \geq 0.25$ VOLTS
$V_{S\ or\ D} \geq 0.30 \leq 0.5$ v
LESS THAN 0.1 OF $V_{S\ or\ D}$ EXCEPT FOR TRANSIENTS AT START AND END OF PERIOD
GATING TIME ±5% of T
TRANSIENTS 10% of T
(B)

Fig. 4. SAGE signals to be delivered to data-using equipment in interconnection unit

A—Timing signal
B—Sync or data signal
Mark Space
Nominal loop impedances = 600 ohms, balanced

serve to amplify the need for knowledge of distribution of errors. To help fulfill this need, laboratory and field programs for error measurement on various types and length of data circuits are being carried on.

In any consideration of data-transmission errors in the SAGE System, a factor of great importance is whether the errors are detected or undetected. An erroneous word which is detected by some device at the receiving terminal of the data-transmission system, can be rejected before it is inserted into the data processing machinery. An undetected word error is fed into the input machinery and consequently has to be processed in order to reveal its false character. In both cases the original word is lost, but in the second case, valuable processing capacity is wasted also. It is then important that error detecting devices be included as part of the transmission and processing systems; some types are already in general use, and others undergoing laboratory and field tests. In general use throughout the SAGE System are simple and interleaved parity checks.

In the simple case, every 16 bits are parity checked; that is, an additional mark or space, as appropriate, is added to the word such that the total of marks adds up to an odd number. Any single false mark or lost mark will disturb the count, and the word can thus be recognized as being in error. Early conjectures indicated that a significant number of errors would occur in adjacent bits in the same transmission. Such errors are normally undetected by simple parity checks. As a protection against such types of consecutive errors in any word, individual words were interleaved.

In an interleaved transmission, the bits in each of a number of words are separated so that the first transmitted sequence consists of all the number 1 bits on n words, the second sequence all the number 2 bits, down to the $n+1$ sequence, which contains all the parity bits. If a, b, c through n are words, and subscripts represent bit order in each word, the sequence is thus:

1st sequence—$a_1, b_1, c_1 \ldots n_1$
2nd sequence—$a_2, b_2, c_2 \ldots n_2$
3rd sequence—$a_3, b_3, c_3 \ldots n_3$
nth sequence—$a_n, b_n, c_n \ldots n_n$
$n+1$ sequence—$a_{\text{parity}}, b_{\text{parity}}, c_{\text{parity}}$ through n_{parity}

Experience will show whether simple or interleaved words are less liable to error. Whether simple or interleaved parity checks are superior in performance is of course a function of the distribution of errors, and until more data are available, no conclusions can be drawn.

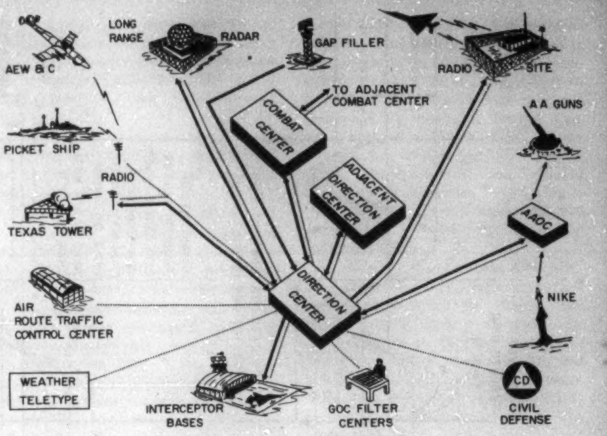


Fig. 5. Typical SAGE sector

—— Data circuits . . . . Other circuits

All data inputs and outputs are parity checked with the sole exception of the gap filler SDV inputs, to which they are not applicable.

System Composition

An elaborate communications network[4] is required to connect the computer[5] to all of the data sources and to the weapons being used for air defense. The basic communication plan developed for the SAGE System combines the technical requirements of the system designer with the operational requirements of the Air Force. One of the most important is optimum reliability at a minimum cost consistent with these requirements. Installations remote from the centralized electronic digital computer that utilize data circuits can be grouped in the following four categories:

1. Surveillance sources (input information).
2. Weapons sites and surveillance control (output information).
3. Similar adjacent computers (input and output cross-telling).
4. Computers at higher Air Defense Command levels (input and output cross-telling).

Fig. 5 illustrates the components of a typical SAGE sector. The heavy interconnecting lines represent the data circuits required between a direction center and other installations. The light lines indicate the remaining voice and teletypewriter circuits that make up the balance of the communication network.

SURVEILLANCE SOURCES (INPUT INFORMATION)

Currently there are five types of surveillance input sources to the centralized computer that employ data circuits. A brief description follows:

1. The largest volume of such input information is received by the computer from a long-range (heavy) radar installation. At a long-range radar site, three kinds of equipment are used to obtain track information. First is the long range surveillance radar. From this, radar returns are obtained from all aircraft within range. Second is the IFF which provides unique beacon assist returns from all friendly aircraft carrying such equipment. This permits tracking aircraft from which there would be an insufficient skin return. Third is the height finder radar. Height finders are necessary because most surveillance radars do not give a target altitude reading. In addition, this unit of equipment provides information that helps in estimating the size of a flight, that is, the number of airplanes in a single flight that are so closely grouped they appear as a single return on a surveillance radar. At this site, information from these three equipments is fed simultaneously into another smaller computer known as the Coordinate Data Transmitting Set. The volume of output information from this set under full load conditions exceeds the capacity of a single 1,300 bit/second data circuit. Consequently, two such circuits are required between this location and the DC.

2. To fill in the low altitude areas not covered by the long range radars, small

gap filler radars are employed. These gap filler radars are automatic, unattended equipments which are connected to the computer by a direct data circuit. This is the only application in which SDV or the 1,600 bit/second data service is employed.

3. Man-made islands known as Texas Towers are anchored to ocean shoals off the Atlantic seaboard. They are so named because of their resemblance to oil derrick platforms used in the Gulf of Mexico. Installations of these towers are similar to those of a long range radar site. Input information to the computer is the same as described in item 1 of this section.

4. Certain long-range radars are so located that their area of surveillance provides coverage not only in the sector in which they are located, but also a considerable and useful coverage in one or more adjacent sectors. When this feature is utilized, the site is called an overlap heavy radar. Since the same considerations apply as when the site functions as only a long range radar site, two additional data circuits are required for direct connection to the input of the computer at each adjacent DC involved.

5. There are two types of guided missile weapons bases that employ data circuits connecting directly to the input of the computer.

    a. The first is the Air Force Bomarc squadron. A single data circuit is employed here to pass missile status information to the DC for automatic insertion into the computer.

    b. The second is from the Army Anti-Aircraft Operations Center (AAOC). This installation controls the NIKE missile. As for the Bomarc squadron, a single data circuit is employed to transmit status information.

These are not the only input sources that are utilized in the system. They are, however, the only ones that currently employ data circuits as direct input information to the centralized computer as shown on Fig. 5. Input information from other sources is transmitted by voice or teletypewriter to a manual input room. At this location, the information is transferred to punch cards, and then fed into the computer. It can be anticipated that the future will see more of the manual input information converted to automatic input information employing some type of data service.

Weapons Sites and Surveillance Control (Output Information)

Besides computers, there are currently four types of data using installations that require output information. Three of these involve control of defense weapons while the fourth deals with surveillance. Following is a brief description:

1. The G/A DL[6] employs frequency division radio equipment. When connected by data circuits from the proper output section of the computer, this equipment will permit the automatic transmittal of flight instructions to manned interceptors. All of the complex co-ordinates and numerical readings required to vector the interceptor pilot to the enemy target are handled automatically. The data link minimized human intervention in routine actions. Because of the design of the DL equipment and the volume of information handled, two data circuits from the output of the computer to the DL ground equipment are required. The principle of interleaving, described earlier, is used to integrate the computer messages for transmission.

2. Control signals for the Air Force Bomarc missile will also be transmitted from G/A DL sites suitably located for this purpose. Although this represents control and flight instructions to a guided missile, rather than to a manned interceptor, the same considerations as discussed in the preceding apply.

3. The inclusion of AAOC in the SAGE complex presented a challenging situation in the integration of an Army-designed weapons system with an Air Force-designed air defense system. The Army Missile Master system for control of the NIKE missile was developed on the basis of receiving data at the rate of 750 bits/second. Since the SAGE computer, however, had not been designed with a 750 bit/second output section, some compromise was necessary. In order to make the two systems compatible, equipment to convert from 1,300 to 750 bit/second rate has been designed.

4. The last type of output data circuit required is to connect the computer to height-finder radars. Here a single circuit is employed for transmitting height request orders to each height-finder in the sector. By this means, height information is obtained on a semiautomatic basis.

As with input information, there are other users of information developed by the computer and personnel within the DC as indicated in Fig. 5. These users do not receive their information automatically from output sections of the computer. Rather this is done manually by voice and teletypewriter circuits. It can be anticipated that in the future more of this information will be handled on an automatic basis. Even at present, transmission over certain teletypewriter circuits is accomplished directly from output sections of the computer rather than by the more conventional means.

Similar Adjacent Computers (Input and Output Cross-Telling

Since an air battle may progress through more than one sector before being brought to a final conclusion, means for handing over control from one DC to another must be provided. For this purpose, automatic cross-telling of tracks between a computer and all adjacent computers is required. A single data circuit is provided for this purpose. Since there is a requirement for information flow in both directions, a total of two such circuits is needed between two adjacent DC.

Computers at Higher Air Defense Command Levels (Input and Output Cross-Telling)

The Combat Center (CC) contains a high-speed electronic digital computer designed to perform its specific functions in the air defense system. From each of the DC computers in the division's area of command, the CC computer receives the sector's air situation summary and combines all of these into an over-all picture for the entire air division. A single data circuit from each DC to the CC is needed for this purpose. With such an over-all summarized air picture available at the CC, command personnel are in a position to arrange for redeployment of defensive weapons as required by the battle situation. Similarly, a data circuit from the CC back to each of its DCs is required.

In all cases, each CC building is constructed adjacent to a DC building. The distance between the two computers is relatively short. The DC computer at such a location transmits the timing, sync and a data sine wave or dipulses directly over house cables to the CC computer.

Just as there is a need for data circuits in both directions of transmission between adjacent DCs, there is a corresponding need between adjacent CCs. Here, two single data circuits (one for each direction) are required to transmit the summary of the air situation and weapon status.

Operational Requirements

In May 1954 Headquarters, Air Defense Command (ADC) made known their operational requirements for restoration of circuits to the Western Electric Company, Air Defense Engineering Service (ADES) Organization. Two different desired restoration times under two conditions of air defense, namely, a "normal" and an "alert status," were presented. Certain critical circuits in a sector required almost immediate restoration while others permitted a slower restoration.

These requirements covered all of the various types of circuits required in the elaborate communications network. The following 3-part discussion will cover only the data circuits and will describe the plan evolved to meet ADC's operational requirements for this group of circuits.

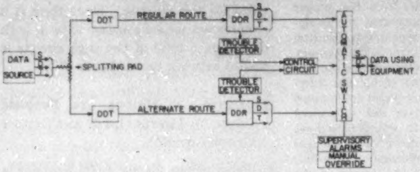
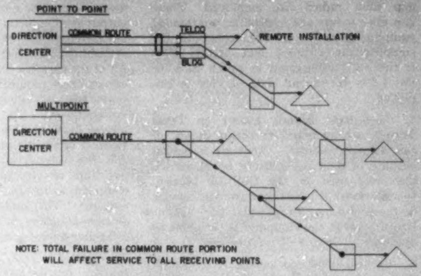


Fig. 6. Dual data facilities

DDT—Digital data transmitter
DDR—Digital data receiver

S—Sync
D—Data
T—Timing



NOTE: TOTAL FAILURE IN COMMON ROUTE PORTION WILL AFFECT SERVICE TO ALL RECEIVING POINTS.

Fig. 7. Point-to-point versus multipoint

Restoration

One of the most important requirements for SAGE circuits is optimum reliability at a minimum cost consistent with ADC's operational requirements. Failure of circuits can occur due to such causes as equipment troubles, human error or acts of nature. In addition, it is essential that the SAGE System capabilities be maintained even under adverse conditions of partial communication failure due to sabotage, bomb damage or other acts of war.

Three methods were considered for the restoration of service on circuits which have failed. The faster the restoration requirement, the more expensive the cost of the method employed.

1. The first method is lowest in cost and longest in time. This involves providing only a single circuit between two points. When a failure occurs, restoration is made by the usual circuit repair methods. Depending on the type of trouble and the particular circuit or facility involved, restoration may be made in minutes or may require hours. Only two types of data circuits rely on this method of restoration.

2. The second method considered was the employment of engineered military circuits (EMCs). By this method, if the single circuit described previously is considered as the regular circuit, where the operational needs are of an extremely vital nature, a second or alternate circuit can be provided. Two such circuits (regular and alternate) are termed a "dual facility." An EMC is not required on a full-time basis, but is established by the commercial communication companies at the request of the military services and is paid for only for the period of time that it is ordered for service. Since the Air Force has been the only user of this grade of data service, the communication companies do not have EMC data-type circuits available as would be true for voice and teletypewriter circuits. Various plans employing an alternate use concept of voice and data were considered that would meet the operational needs of the Air Force and still provide the economy of EMCs. Because of the operational complexity of even the simplest of these plans, the ADC reluctantly decided that this method was not feasible for data circuits. However, EMCs are utilized for the other types of communication services.

3. The third method is basically the same as the one just described, except that the alternate circuit is ordered on a full-time basis as is the regular circuit. When such a dual facility is provided, a decision is necessary to determine whether the transfer of service from the regular (when failure occurs) to the alternate should be accomplished manually or automatically. Operational requirements, together with the number of circuits involved, dictated that this be done automatically. Since a data circuit is a one-way transmitting facility, the problem is simplified. By parallel or bridge-feeding the sending end of both circuits, each circuit will be energized simultaneously with the same information. Fig. 6 shows the requirements established to meet this method of operation. A trouble detector is located at the receiving end of each of the two circuits. Normally the data using equipment takes service from the regular circuit. When trouble occurs on this circuit, the detector actuates the switch or relay to transfer to the alternate circuit. In the event both circuits are in trouble simultaneously, no transfer occurs. Visual and audible alarms are specified as requirements along with a manual control to provide an override feature.

Leased communication facilities are provided on a dual basis consistent with other components of the system.

Route Selection

Where dual facilities are specified, it would be unwise to provide both the regular and alternate circuits over the same geographic route and type of facility. Dual circuits on the same route would offer little or no protection against the ravages of nature or acts of war. Even without the provision of dual circuits, a measure of protection against outages from any cause would be achieved by dividing regular circuits among two or more routes.

Because of these considerations, where dual circuits are specified, at least two routes are also included as part of the requirement. These are to be completely separated. Routes have been selected which avoid major target areas.

In general, no attempt is made to specify the basic type or types of facility that should be utilized on each of the two routes. The communication companies are aware of the fact, however, that a diversification by types of facilities between routes is desirable in cases where this will improve reliability. As an example, underground facilities are less likely to suffer damage from a hurricane than say an open wire or an aerial cable line. On the other hand microwave radio relay facilities may be easier to protect against acts of sabotage than the open wire or aerial cable type of telephone plant.

Special Cases

For a majority of the types of remote locations that receive output information from the SAGE computer, the full 1,300-bit capacity of the data circuit is not required, or the same information is used at more than one location. This permits using a single computer output section to feed information to more than one location. If different messages are required at different locations, appropriate addresses are inserted in the messages to permit the required location to accept only the needed information. In these cases, a number of points such as adjacent DCs are fed from this one output section. Economies in charges for such leased circuits can be achieved if, instead of ordering individual point-to-point circuits from the transmitting DC, a multipoint circuit is ordered.

In addition, data circuits from a long range radar that serves as an overlap heavy radar are provided to the parent

DC as well as to adjacent DCs on a multipoint basis.

Where data circuits are ordered on a multipoint basis, a requirement has been stipulated that the bridging arrangements provided be such that trouble on one leg of the bridge cannot feed back into other working legs. Even with this safeguard, one might question whether employing a multipoint arrangement was logical with all of the emphasis that has been placed on reliability and circuit restoration. If the first bridge point were 50 miles from the originating DC, any failure in this first 50-mile section would affect the service to all receiving points.

However, in the common route portion of circuit facilities, the point-to-point arrangement is just as vulnerable to failure due to acts of war as the multipoint arrangement. In some cases, the same is true with regard to outages occasioned by day-to-day maintenance troubles, as in Fig. 7.

Development, engineering, manufacture, installation, test, and evaluation are all going on simultaneously due to the urgency of the SAGE program. Advancement in the state-of-the-art imposes changing requirements on the data circuits. A case in point is that of G/A DL communications.

The frequency division G/A DL system, previously mentioned, has been designated by the Air Force as an interim system, and is to be replaced by a time division G/A DL system. In the frequency division system, every ground radio transmitter in a sector is simultaneously transmitting the data generated by the computer. In the time division system, however, only one transmitter will be radiating at any given time.

Technical requirements of this new system are such that a new limit will be imposed on the data circuits which will feed the time division G/A DL equipment. Because of the composition of the message format and to avoid overlapping as the radio transmitters are keyed sequentially, there is a requirement that computer-generated messages should arrive at all sites over all circuits and routes almost simultaneously. Since the different circuits will have varying lengths and will consist of different kinds of facilities with differing velocities of propagation, there will be a variation in transmission speeds. A 2-millisecond differential limit has been established for these circuits. As a consequence, differential delay equalization will be needed to meet this 2-millisecond requirement. Multipointing is feasible for both the frequency and time division system.

Conclusions

While considerable progress has been made in establishing and meeting the needs for SAGE data system requirements, it should not be assumed that all work has been completed in this field. The following conclusions are reached:

1. Data transmission requirements for SAGE have been prescribed. These combine the technical requirements of the system designer with the operational requirements of the Air Force. The important requirement of optimum reliability at a minimum cost has been carefully considered. A workable method of meeting these requirements is available.

2. More experience is needed to determine ultimate data transmission standards for optimum bit rates and error rate per channel. Refinements are needed since the present data system integration is not optimum in this regard as indicated by the difference in bit rate for SAGE and an AAOC.

3. A continuing conversion of inputs and outputs from manual to automatic is to be expected.

4. New surveillance and weapons systems can be expected which will need to be integrated with SAGE. The data transmission requirements for these new systems are not yet known.

5. Improved techniques arising from this military application will find application also for industrial users.

Appendix I. Data Circuit Characteristics

1. *General.* The circuit shall be suitable for unidirectional transmission only.

2. *Impedance.* The circuit shall be designed to work into terminal impedances of 600 ohms, resistive.

3. *Net Loss.* The 1,000-cycle net loss of the circuit shall be 20 db (decibels) with not more than ±1.0 db daily variation and ±4.0 db seasonal variation.

   a. The net loss in the preceding is effective between the customers' sending and receiving equipment. For circuit layout purposes normal telephotograph practices will apply.

4. *Frequency Response.* The circuit shall transmit all frequencies between 500 and 2,500 cycles with not more than ±3 db variation over this band. Between 300 and 500 cycles the response shall not exceed the 500-cycle response by more than 2 db. Between 2,500 and 3,000 cycles the response shall not exceed the 2,500-cycle response by more than 2 db.

5. *Envelope Delay.* The total difference in envelope delay between any two frequencies in the band from 1,000 to 2,500 cycles shall not exceed 500 microseconds (μsec).

   a. On data circuits used for multiplexed transmission in groups of two or three, and specifically ordered as such, the difference in envelope delay at 2,000 cycles between any two circuits in the group shall not exceed 300 μsec, with not more than +50 μsec variation from this figure in long-time variation.

6. *Noise.* The weighted noise level, measured as below, shall not exceed 40 db below the synchronizing signal level. The number of impulse noise peaks which equal or exceed 18 db below the synchronizing signal level shall not exceed 1 per minute.

   a. The noise weighting characteristic used above shall be the standard *F1A* plus the telephotograph network, corresponding to present telephotograph practices.

   b. The data circuits will be terminated in well-balanced coils and all normal precautions will be observed as to circuit unbalances, etc. The customers' equipment shall also include a well-balanced input transformer with no connections to ground on the line side of the transformer.

7. *Modulation.* The voltage sum of the modulation components of the signal generated in the data circuit due to 60, 120, and 180 cycles shall not exceed 30 db below the synchronizing signal level.

8. *Transmitting Level.* The maximum signal level fed to the data circuit at the transmitting end shall not exceed the peak-to-peak voltage of 0 dbm into 600 ohms. In accordance with item 3 of the foregoing, the signal voltage at the receiving end of the data circuit will be 20 db below this.

Appendix II. Signal Characteristics

Four kinds of data are transmitted as follows: a sync pulse, a frame pulse, timing pulses, and data pulses. The modulation is applied as a square wave of varying amplitude (see Fig. 1) on a carrier of 2,000 cycles, or of 1,950 cycles.

If any one piece of data is defined as a bit, then the bit rate is 1,600 pulses per second or 1,300 pulses per second or the duration of a bit is 1/1,600 second (625 μsec) or 1/1,300 second.

The form of a typical signal with relative amplitudes and tolerances is shown in Fig. 2.

It will be seen that if the sync pulse has a peak amplitude of 1.0, then the unmodulated carrier has an amplitude of 0.25. The carrier plus superimposed timing has an amplitude of 0.31. The carrier plus a data pulse has an amplitude of 0.63, and the carrier plus data plus timing has an amplitude of 0.69. The frame sync pulse has an amplitude of 1.0 and occurs in the time interval immediately following the sync pulse.

A signal which is greater than 0.44 in amplitude is construed as a data pulse. A signal greater in amplitude than 0.75 is construed as a sync or frame sync pulse.

References

1. SAGE Provides New Defense Concept, Philip J. Klass. *Aviation Week*, New York, N. Y., vol. 64, no. 5, Jan. 30, 1956, pp. 46–55.

2. SAGE Data Terminals, R. O. Soffel, E. G. Spack. *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 872–79.

3. Communication Channels for SAGE Data Systems, R. T. James. *Ibid.*, pp. 838–43.

4. SAGE—Our New Air Defense System, Gordon A. Blake. *Signal*, Washington, D. C., vol. 10, no 3, Jan.-Feb. 1956, pp. 15–20.

5. SAGE Air Defense Computer, John M. Carrol. *Electronics*, New York, N. Y., vol. 29, no. 10, Oct. 1956, pp. 234, 236, 238.

6. Data Link May Start Vortac-Type Battle, Philip J. Klass. *Aviation Week*, vol. 66, no. 13, April 1, 1957, p. 87.

Synchronized Clocks for Data Transmission

J. O. EDSON
NONMEMBER AIEE

M. A. FLAVIN
NONMEMBER AIEE

A. D. PERRY
NONMEMBER AIEE

Synopsis. Accurate timing waves for the transmission and decoding of a kilobit per second data signal are generated by the use of tuning fork resonators in a flywheel-type of synchronization circuit. Accuracy of the derived timing wave is good enough that the timing jitter makes a negligible increase in the error rate. Synchronization is established in about 20 milliseconds prior to sending data.

DIGITAL data transmission depends upon the transmission of time variable electric waves. Usually, the signal takes the form of a succession of pulses of prearranged duration. The data are recovered with the aid of a clock of some sort. If the transmission speed is very low compared to the bandwidth of the medium, the clock may be so simple that its presence is not recognized. If maximum speed is to be attained, transmitting and receiving clocks must maintain synchronism with error negligible compared to the duration of a pulse.

A clock is a device which marks periods of time which should all be equal. The operation of a clock can be specified by the period T of its marks and by a smaller time t_0 specifying the displacement of one mark from a reference, or zero, time. Two clocks, standing side by side, are synchronous if these two parameters for one clock are the same as those for the second. A transmitting clock controls the length and spacing of information bearing pulses. The minimum spacing of pulses is a useful unit of time and will be called a "tick." A receiving clock is used to analyze or decode the received sequence of pulses. When these clocks are separated by a significant distance, the idea of synchronism needs to be extended slightly. The electric signal is delayed in transmission. Usually, there is some dispersion or phase distortion in transmission. It is convenient to consider the clocks synchronous when the receiving clock has the same period T as has the transmitting clock but lags by a time equal to the time of flight or envelope delay of the most significant portions of the message wave. This relation can be established by sending and receiving some prearranged starting signal. This operation permits setting the zero of the receiving clock to agree, within some tolerance, with that of the transmitting clock. When this is done, the delay in transmission is not important provided it does not change after the starting signal is sent.

The teletypewriter code is sent by the aid of a simple timing method: start-stop synchronization. The advantages of this well-known system are simplicity and capability of sending short messages (5 bits long) at irregular intervals. Its disadvantages are that it uses about 30% of the time on the transmission line for establishing synchronism, and that each "setting" of the receiving clock depends on a single transition from marking to spacing and so may suffer substantial error.

If the typical message is much longer than 5 bits, a starting signal can last for many ticks without adding greatly to the total transmission time. The error in setting the receiving clock tends to vary inversely with the duration of the starting signal, which is usually a sequence of alternate marks and spaces. The message wave also contains some information about the transmitting clock. This can be used to improve the zero setting and to adjust the period of the receiving clock.

Requisite of a Good Synchronization System

A good system of synchronization will use a minimum time before the beginning of a message for setting the receiving clock to agree with the transmitting clock. It will be able to bridge any probable run of marks or spaces in transmission. It will not depend upon any one or small number of transitions for establishing its timing, but will average the effects of many.

In transmission of random data, there is no restriction on the length of run of marks or spaces but a practical upper limit is about 50. However, some plans envision back and forth receive and confirm operation. In this case, the receiving clock should remain in step for about one second with no received signal.

Methods of Synchronization

Two general methods of synchronization exist. An oscillator having nearly the correct frequency can be used as the receiving clock. The zero set can be accomplished by adjusting a phase shifter or delay line under control of the start signal. An automatic frequency control system can be set up to compare the time of pulses from the receiving clock with the time of receiving corresponding pulses from the transmitter and to adjust the frequency of the receiving clock to minimize the average discrepancy. Alternatively, a narrow-band filter or resonator may be driven by pulses derived from the received wave to serve as the receiving clock. This second system is often called flywheel synchronization. Either method tends to average the effects of many received message pulses in order to determine suitable values of period T and zero set t_0. It is believed that the controlled oscillator should give superior results, but the flywheel system is very good and is simpler.

FLYWHEEL SYNCHRONIZATION

If the information is transmitted by rectangular pulses of duration T sent or

---

Paper 58-300, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted April 3, 1958; made available for printing July 24, 1958.

J. O. Edson and M. A. Flavin are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J., and A. D. Perry, formerly with Bell Telephone Laboratories, Inc., is now with the General Electric Company, Ithaca, N. Y.

suppressed at regular intervals T, then the spectrum of these pulses contains no lines at the bit rate or its harmonics. Limiting the received signal produces a rectangular wave which tends to be positive for one half the time as does noise. It differs from noise in that the transitions from plus to minus or from minus to plus occur only at certain permitted times at intervals T. A differentiating network can produce short positive and negative pulses at these transition instants. The spectrum of either the positive or the negative pulses comprises a noise-like continuum and lines at $1/T$ and its harmonics. A full-wave rectifier serves to invert the negative or positive pulses, depending on design, and to accentuate the lines compared to the noise-like background. A high Q resonator can select the fundamental line or a harmonic and further suppress the background noise. If the resonator operates on a harmonic of $1/T$, it must be followed by a divider chain which produces the bit frequency.

When the message has been transmitted over a medium having minimum bandwidth and some delay distortion, the pulses from the differentiator tend to be advanced or retarded relative to the instants at which they would otherwise have occurred. Line noise aggravates this effect. This results in phase modulation of the frequency $1/T$. A high Q resonator averages the effects of many pulses and thus reduces the phase modulation. In addition, it can ring for many periods without excitation, bridging long runs with no transitions. The disadvantages of high selectivity are that better frequency accuracy is required and that it is more difficult to build up the oscillation quickly at the beginning of a message. The start problem has been solved by driving the resonator with pulses of about eight times normal amplitude for about 20 milliseconds. This brings the response up to a sufficient amplitude to supply satisfactory timing pulses.

CHOICE OF A RESONATOR

The resonators in a flywheel synchronization system must meet a number of severe requirements. The receiving resonator is particularly critical. It must have high Q, good frequency stability, a large ratio of bandwidth to drift of resonant frequency, and small undesired transmission. A resonator meeting the receiver requirements will be adequate for the transmitter as well.

The need for high Q has been mentioned in the previous section. A Q of several hundred to several thousand will provide both sufficient averaging time and adequate performance in bridging gaps of several hundred to several thousand ticks.

During a gap the receiving resonator rings at its natural frequency. If this frequency differs from the selected harmonic of the transmitting bit rate, timing errors will accumulate. If the error at the end of a 100-tick gap is to be less than 1/10 tick (which is not excellent performance), the frequency error may be no more than 0.1%. Assigning half the tolerance to the receiver gives accuracy requirement of 0.05%, or 500 parts per million (ppm). For 1/10 tick error at the end of a 1,000-tick gap the tolerance is 50 ppm.

Except during gaps, it is convenient to think of the receiving resonator as selecting the line in the spectrum of its input signal at the nth harmonic of the bit rate and rejecting the remainder of the spectrum, particularly the noise components. Any deviation of the resonator frequency f_r from the frequency of this line nf_b introduces a phase displacement in the resonator output and a corresponding phase change in the recovered timing wave. If the system is initially set on tune and then drifts, the phase error in the recovered timing wave, for small drifts, is:

$$\Delta\varphi = \frac{1}{n} \tan^{-1} 2Q\left(\frac{\Delta f_b}{f_b} - \frac{\Delta f_r}{f_r}\right)$$

The possibility of limiting this error by increasing n is limited by the presence of jitter in the received transitions. The resonator measures jitter in units of its natural period, and thus sees an effective jitter which increases directly with n. Since the strength of the desired line in the resonator input spectrum falls off and the noise level increases, as the effective jitter increases, operation at large n carries a signal-to-noise penalty. Values of n greater than one put a premium on good signal-to-noise ratio because a missed or added zero-crossing in the resonator output throws the divider chain out of step. Finally, large values of n lead to difficulty in getting the divider chain in step initially.

The disadvantages of large n are great enough that $n = 1$ would be the preferred value if suitable resonators were available. Values greater than two impose impractically severe limitations on jitter. If we require that detuning errors introduce a timing error of at most a tenth of an interval, the ratio of bandwidth to frequency error (for each resonator) must be greater than 13 (or 6.8) for $n = 1$ (or 2). With a Q of 500, the corresponding frequency stabilities are 155 and 300 ppm.

Q's of several hundred are not easily obtained with passive inductance-capacitance circuits in the 1 kilocycle region. While it is possible that active Q-multiplier circuits could be used, stabilities of a few hundred parts per million are difficult enough to achieve with passive elements that the added drifts of active elements make the approach unattractive.

One may achieve excellent stability with resonators involving mechanical motion—tuning forks and crystals. With these elements the frequency accuracy is certainly sufficient for bridging 100-tick gaps. The Q's, however, will be 5 to 10,000 for forks and 50,000 or more for crystals. Even with $n = 2$ (which was selected for the experimental models) required stabilities are of the order of 20 ppm for forks and 3 ppm or less for crystals. Since these tolerances are severe, it is desirable to look for ways around them.

Two alternate approaches are available. One may reduce the Q of the circuit by applying negative feedback around the resonator, or he may increase the bandwidth by translating the input signal up in frequency and the output down by modulation techniques. Both techniques have been used in the past in less critical applications. Negative feedback raises the level of undesired transmission and runs into trouble with secondary resonances of the vibrating system. Modulation techniques demand the matching of two resonators so that the difference of their frequencies is the desired transmission frequency. Neither method is simple.

A single thermostated precision audio-frequency quartz crystal might provide an adequate ratio of bandwidth to frequency drift. It would, however, be bulky, expensive, and subject to breakage in normal handling. At least the thermostating and handling problems may be eliminated by using a precision tuning fork.

Because the most common application of tuning forks is in frequency control of oscillators, not all manufacturers supply units suitable for use as filters. There are two common defects: secondary resonances which introduce undesired peaks in the transmission curves, and insufficient isolation between the input and output, which produces an undesirably large nonresonant coupling. Forks supplied by Times Facsimile Corporation have been found free of spurious transmission peaks. The standard models provide about 30 decibels more resonant than nonresonant transmission, which is usable but definitely less than one would like.

These forks have a temperature coeffi-

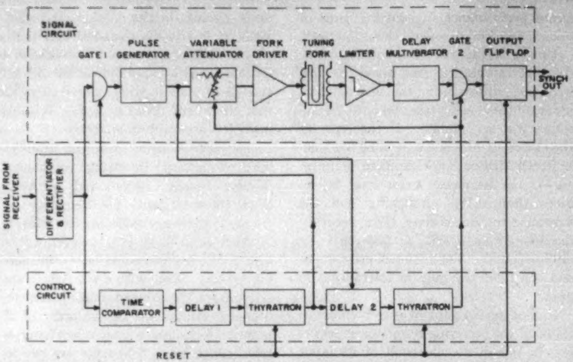


Fig. 1. Synchronization recovery system

cient of less than 0.2 ppm/degree centigrade in the 10–65 degrees centigrade range, which is acceptable. The aging rate claimed is less than 5 ppm per year. This rate is acceptable for experimental work. In field service, aging would be corrected by periodic retuning, which is accomplished by adjusting the drive and hence the amplitude of oscillation. This in turn changes the resonant frequency, by about 75-90 ppm for a change in drive from one volt to 5 volts. This nonlinearity in fork characteristics is useful in the transmitter since it provides a simple means of adjusting the clock frequency, but in the receiver it leads to a variation in output phase when the average number of received transitions per second varies. Because of the long averaging time, this effect is of no consequence with random data signals, but it is a source of error when a single word is repeated many times.

Further improvement in the resonator part of the system is readily possible. Operating with $n = 1$ will reduce the importance of nonresonant transmission, and forks now under development will reduce its level. Since the drive-frequency dependence of the fork is less at low drive levels than at high, grinding the forks to correct frequency with small drive will reduce the dependence of output phase on the number of transitions in the data signal. Still more may be done, but as one seeks still better performance, the cost rises rapidly.

Synchronization Recovery System

A synchronization recovery system has been designed which takes advantage of the good stability and long memory of the tuning fork. The system makes a good estimate of phase because the fork integrates over a large number of transitions. The system is started quickly by initially overdriving the fork. The initial estimate of phase is made on the basis of a few transitions but tends to be accurate because of the constraints on the signal. During the start period the signal is required to be rhythmic mark-space and the fork drive is inhibited until the mark to space ratio nears unity.

The system which has been built uses a 1,600 cps (cycles per second) tuning fork. It can provide timing for either an 800 or 1,600 bit per second signal. For 800 bits per second the fork operates on the second harmonic of its input, the fork output frequency is divided by two. The division is controlled so that only one of the two possible output phases is allowed.

SYSTEM BLOCK DIAGRAM

The block diagram is shown in Fig. 1. The arrangement for 800-bit-per-second operation is shown since this is the more complex case. The signal input to the synch recovery circuit is a binary wave, reconstructed in amplitude but jittered in time. This wave is differentiated and full-wave rectified to produce positive pulses marking the transitions of the input wave. These pulses are applied to the signal circut and the control circuit.

SIGNAL CIRCUIT

The pulses marking the transitions are applied to a gate which is enabled when the control circuit indicates that the synchronizer input has acceptably low distortion. The output of this gate drives a single-shot multivibrator which generates pulses whose length is a half period of the oscillation of the fork. These pulses are passed through an attenuator to the fork driver which applies them to the tuning fork itself. The attenuator is initially set for low loss to provide high drive for the fork during the start-up period. When the control circuit supplies a signal indicating that adequate time has been provided for the build up of the fork, the attenuator is switched to a much higher loss. This loss reduces the drive to that which is necessary to maintain the amplitude of fork oscillation at the desired level.

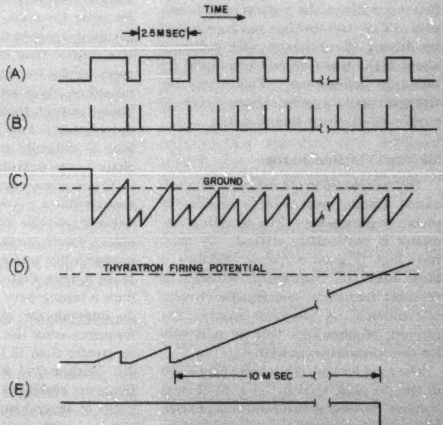


Fig. 2. Waveform diagram illustrating control of thyratron 1

A—Input from receiver
B—Differentiated and rectified
C—Time comparison
D—Delay 1
E—Plate of thyratron 1

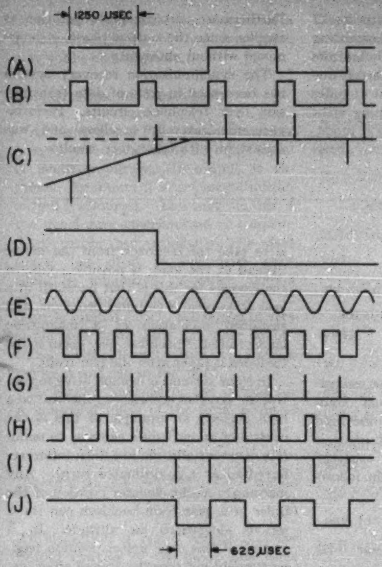


Fig. 3. Waveform diagram illustrating control of thyratron 2

A—Signal in
B—Pulse generator output
C—Thyratron 2 grid
D—Thyratron 2 plate
E—Fork out
F—Limiter out
G—Delay multivibrator in
H—Delay multivibrator out
I—Flip-flop in
J—Flip-flop output

The output voltage of the fork is passed through a limiting amplifier which converts it to a rectangular wave. The rectangular wave is differentiated, generating pulses to drive a delay multivibrator which in turn drives the output flip-flop. The delay multivibrator corrects for accumulated phase shifts in the recovery circuit and is made adjustable to permit matching the transitions in the output with the transitions in an unjittered signal. The output flip-flop counts down by two from the frequency of the fork and provides a square-wave output in which both transitions are accurately timed. The two outputs of the flip-flop are applied to isolating amplifiers which in turn drive the connected loads.

CONTROL CIRCUIT

Pulses marking the transitions are fed into the time comparator circuit. This circuit is an integrator which responds to the pulses as shown in Fig. 2(C). Whenever the time comparator output exceeds ground, it causes the resetting of the delay, see Fig. 2(D). When the mark and space periods approach equality, the time comparator output no longer exceeds ground, and the delay is no longer reset. The delay is another intergrator which charges about eight pulse periods before reaching a thyratron firing voltage. When the thyratron fires the circuit knows that the mark and space periods have become about equal and have stayed that way for several periods. It is assumed, therefore, that severe distortion is ended. When the first thyratron fires it enables gate 1 and starts a delay. After about 20 milliseconds the second thyratron fires, reducing the drive on the fork and enabling gate 2 to apply pulses to the output flip-flop.

The firing of the second thyatron must be synchronized with the input signal to insure that no ambiguity occurs in the output. Wave shapes of Fig. 3 indicate how that is accomplished. The pulse generator of the signal circuit generates 312.5-microsecond pulses; see Fig. 3(B). This wave is differentiated so that the positive pulses correspond to the delayed transitions. The differentiated pulses are fed to the thyratron grid where they are superimposed on the charging curve of the delay integrator; see Fig. 3(C). The second thyratron fires after a delay of about 20 milliseconds and at the occurrence of a positive pulse. Note that the voltage change between pulses is less than the amplitude of a positive pulse, assuring that the thyratron will actually fire on a pulse. The first pulse out of the delay multivibrator after the thyratron fires is gated to the output flip-flop; see Fig. 3(I). This mode of operation allows 312.5 microseconds of jitter on the input wave without incorrect starting. With the relative phases and voltages shown in Fig. 3, the output flip-flop is constrained to start with a negative going transition at the bit center.

Results

The timing of the output of the synchronization recovery system is slightly affected by the frequency drift of the transmitting oscillator and by the input code. The output transitions vary about 6 microseconds relative to the input transitions for a frequency drift of 7 ppm. The drift will be less than 7 ppm if the transmitter is tuned periodically. The input code to the system is expected to be nearly random, which implies that on the average, half the possible transitions will be present. If the timing is accurate when 50% of the transitions are present, an error of 6 microseconds will occur when a code with only 40% of the transitions present is applied continuously. However, it is highly unlikely that only 40% of the possible transitions will occur for a time comparable to the memory of the fork if the code is nearly random. Therefore, the error due to the variation in the code is expected to be very small.

Tests were run on the effect of jitter. Fig. 4 shows the relation between the jitter of the recovered timing wave and the jitter of the input wave. The figure shows that the jitter has less effect at the 1,600 bit rate. When the input jitter exceeds 25%, the output flip-flop will sporadically miss a count. This will cause the 800-bit-per-second output to jump 180 degrees. This effect does not occur

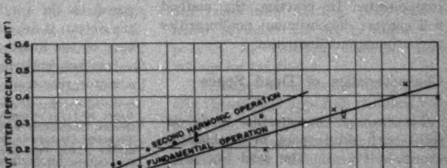


Fig. 4. Jitter characteristic of synchronization recovery circuit. For this experiment input and output jitter distributions were approximately Gaussian

at 1,600 bits since no frequency division is used at this bit rate.

Conclusions

The synchronization recovery circuit which has been described provides an accurate timing wave in the presence of transmitting oscillator drift, variation of codes, and jitter. If the transmitting oscillator is adjusted periodically and the data are nearly random, timing errors due to oscillator drift and variation of codes will be small. The major timing error will be due to jitter. Operation at fundamental has greater tolerance to jitter. Furthermore, fundamental operation is simpler since the output phase is determined without ambiguity.

The synchronization recovery system has been used in tests of data transmission over telephone circuits. Performance was generally excellent and was consistent with laboratory results.

Linearization of Dead Space

SHELDON JONES
ASSOCIATE MEMBER AIEE

Synopsis: The characteristics of dead space are reviewed, and a method of converting the discontinuous nonlinearities of dead space to a linear characteristic or, at least to a continuous nonlinearity, is outlined.

DESIGNERS of most physical systems have a wealth of information available for the synthesis of linear systems. Of course, the reaction of a true system to small perturbations will not be precisely as predicted because of the linear approximation of the system differential equations, i.e., assuming constant coefficients from one signal level of operation to another. Fortunately, the linearization of continuous nonlinearities, in general, does not create serious deviations from the stability prediction of linear analysis.

Discontinuous nonlinearities such as static friction, hysteresis, and saturation exist in most physical systems. Extensive theories have been developed for the analysis of this type of nonlinearity. The most common of these theories are those of the phase plane and the describing function. Unfortunately, these theories do not contribute substantially to the synthesis of physical systems of any complexity.

A method is outlined for linearization of dead space or backlash with ideal components. In practice, the method will convert discontinuous nonlinearities to continuous nonlinearities.

Characteristics of Dead Space

The transfer characteristic of an element with dead space or backlash is shown in Fig. 1. The solid line is an idealized plot of output versus input for an element with this characteristic. The dashed portion of the transfer characteristic represents more nearly the output-input relationship with actual components. The output of the idealized device to a specific sinusoidal input is plotted in Fig. 2. A Fourier analysis[1] of the output waveform yields the following equations:

$$y(t)=C_1 \sin \omega t+\sum_{n=3,5\ldots}^{\infty} C_n \sin n\omega t \tag{1}$$

where

$$C_1=\frac{EK}{\pi}(\pi-2B+\sin 2B)-\frac{4Kb \cos B}{\pi} \tag{1A}$$

$$C_n=\frac{4K}{\pi}\left[\frac{E}{1-n^2}(\sin nB \cos B - \eta \sin B \cos nB)-\frac{b}{n}\cos nB\right] \tag{1B}$$

$$B=\sin^{-1}\frac{b}{E} \tag{1C}$$

An input to a system of this type must exceed a certain magnitude before an output is obtained. When the signal amplitude is twice the dead space, the third harmonic amplitude is approximately 29% of the fundamental frequency amplitude. When the signal amplitude is 1.156 times the dead space, the third harmonic is 67% of the fundamental. For large values of threshold the harmonic content of the output waveform becomes appreciable when compared to the fundamental. Of course, the output is dependent upon the amplitude of the input signal.

One frequently used design criterion for a motor gear train driven position servo is to take the feedback from the motor instead of the load, if possible. One of the reasons for this is that backlash in a gear train creates stability problems, and sustained oscillation of a magnitude larger than the backlash may occur if the feedback is taken after the gear train.

In some systems is is imperative to control the position accuracy of the load to a high degree; an example of this is the control of an airplane's elevator to maintain constant altitude in a turn (often referred to as a co-ordinated turn). Any inaccuracy in the elevator position of the order of a gear train backlash can cause serious deviations in altitude. In a system of this type motor position feedback might not give the desired accuracy.

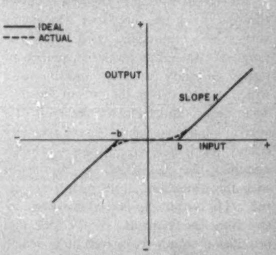


Fig. 1. Output-input relationship for a device with dead space or backlash

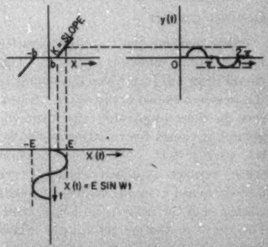


Fig. 2. Effect of dead space on output waveshape

Paper 58-1174, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 27, 1958; made available for printing August 4, 1958.

SHELDON JONES is with the Hughes Aircraft Company, Culver City, Calif.

Of course, in certain types of systems a sustained oscillation of small magnitude may not be objectionable.

Linearization of Deadspace

An inspection of Fig. 2 shows that, over a region corresponding to an input signal of small amplitude, no output is obtained; for inputs of larger magnitude an output is obtained. This suggests that if the signal were superimposed on a square wave carrier proportional to the dead space, the effect of the dead space could be minimized. Another method of accomplishing the same thing is to provide infinite gain in the region of zero signal and finite gain for other amplitudes. The characteristics of such a device are shown in Fig. 3(A).

A Fourier analysis of the output for a sinusoidal input of the form $E \sin \omega t$ is

$$f(t) = EK \sin \omega t + \sum_{n=1,3,5,7\ldots}^{\infty} \frac{4a}{n\pi} \sin n\omega t \tag{2}$$

An inspection of Fig. 3 reveals that the output consists of the input sine wave plus a square wave.

Fig. 4 depicts a system where a nonlinearity of the form shown in Fig. 3 is connected in tandem to the nonlinearity form shown in Fig. 1.

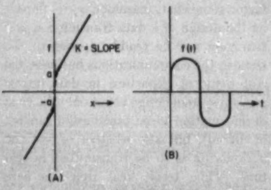


Fig. 3. Input-output characteristic of nonlinear amplifier

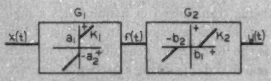


Fig. 4. Block diagram of dead space and nonlinear driver

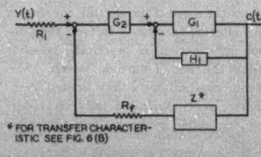


Fig. 5. Functional diagram of high-gain nonlinear amplifier

The input-output relations for block G_1 are

$$f(t) = K_1 x(t) + a_1;\ x(t) > 0 \tag{3}$$

and

$$f(t) = K_1 x(t) - a_2;\ x(t) < 0 \tag{4}$$

The equations for block G_2 are

$$y(t) = K_2[f(t) - b_1];\ f(t) \geq b_1 \tag{5}$$

and

$$y(t) = K_2[f(t) + b_2];\ f(t) \leq b_2 \tag{6}$$

Substitution of equation 3 in equation 5 yields

$$y(t) = K_2[K_1 x(t) + a_1 - b_1];\ x(t) > 0 \tag{7}$$

and for $a_1 = b_1$

$$y(t) = K_2 K_1 x(t);\ x(t) > 0 \tag{8}$$

Similarly, substitution of equation 6 gives

$$y(t) = K_2[K_1 x(t) - a_2 + b_2] \tag{9}$$

and, for $a_2 = b_2$,

$$y(t) = K_2 K_1 x(t);\ x(t) > 0 \tag{10}$$

An inspection of equations 3 and 4 shows that for $t = 0$, $f(t)$ will be $+a$ or $-a$. Therefore, it is concluded that

$$y(t) = K_2 K_1 x(t) \tag{11}$$

for all inputs to the system.

Construction of Nonlinear Amplifier

The input-output relationship exhibited in Fig. 3(A) can be achieved by building an amplifier with infinite open-loop gain and operating the amplifier open-loop closed-loop as a function of output voltage. That is, feedback could be applied to the amplifier at a certain magnitude of output voltage. Fig. 5 is a block diagram of such a system. The inner loop transfer function G_1H_1 can be made an unstable loop. This would approach infinite gain with the outer loop open. The outer loop can be stabilized for values of output voltage greater than the dead space exhibited by the feedback element Z. Ideally, the feedback Z would offer infinite impedance up to a certain level and zero impedance above this level of output. In addition, it might be desirable to make the insensitive zone of Z asymmetrical, in the event of asymmetrical dead zone in the load.

A device exhibiting the characteristic of the feedback element Z is shown in Fig. 6(A). The device consists of two Zener diodes connected "back to back," i.e., either the anodes or cathodes tied together.

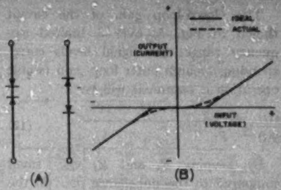


Fig. 6. Nonlinear feedback element

A—Back-to-back connection for Zener diodes
B—Output-input relation of A

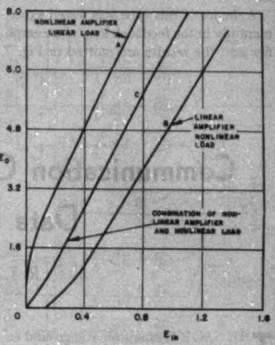


Fig. 7. Experimental data

A Zener diode exhibits the property that very little current flows when reverse voltage is applied to the diode until the breakdown or Zener voltage is reached. For voltages applied in excess of the Zener voltage the diode is equivalent to a resistor and a battery in series. When two diodes are tied together, as shown in Fig. 6(A), the input-output relationship is as shown in (B), where current is considered the output and voltage is considered the input. The input-output relationship of Fig. 6(B) can be made asymmetrical as a function of polarity by choosing different Zener voltage for the diodes.

The Zener diode combination, if properly chosen, will approach the ideal characteristic for the feedback element Z of Fig. 6(B) (see dashed and solid line). Actually, this "nonideal" characteristic may be desirable since in practice dead space exhibits the same characteristic (see dashed line of Fig. 1).

The closed-loop gain of the circuit shown in Fig. 5, over a limited frequency range, for signal levels corresponding to high outer loop gain (stable operation is assumed) will be

$$\frac{c(t)}{y(t)} = \frac{R_f+Z}{R_1} \tag{12}$$

By keeping R_1 and R_f very small compared to the impedance of Z in the "conducting region," the amplifier will appear to be operating without feedback; i.e., the feedback resistance will be infinite. In the conducting region the amplifier gain will be given by equation 12.

A device of this type was tested experimentally in the feedback loop of an amplifier and the results are plotted in Fig. 7. The amplifier, a low-gain amplifier, was chosen because of availablity.

Fig. 7, curve B, is a plot of output versus input when the amplifier was operated as a linear amplifier driving a nonlinear load. With the Zener diodes in the feedback loop, the input-output relationship of the nonlinear amplifier driving a linear load is shown in curve A, Fig. 7. When the nonlinear amplifier was used to drive a device with dead space, the input output relation of curve C, Fig. 7 was obtained.

Conclusions

In a system with dead space, the use of a nonlinear amplifier such as has been described yields the following advantages:

1. The stability of the system is improved and the possibility of low-level oscillations is minimized.
2. The sensitivity in the low-signal region is increased.
3. It tends to linearize the input-output relationship of the open-loop system.

The conversion of the discontinuous type of nonlinearity to a continuous nonlinearity allows the designer to base stability prediction on linear analyses. In addition, the designer of position feedback systems has more latitude in choosing the output feedback point.

Reference

1. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN, VOL. II (book), Harold Chestnut, Robert W. Mayer. John Wiley & Sons, Inc., New York, N. Y., 1955.

Communication Channels for SAGE Data Systems

R. T. JAMES
NONMEMBER AIEE

THE SAGE (semiautomatic ground environment) System is a system of air surveillance and control designed to aid in the air defense of the continental United States. In this system of air surveillance, flight information is transmitted from radars and other sources to a central point where it is processed and evaluated to form a comprehensive picture of the air situation. This picture covers the search area of several radars, thus providing centralized surveillance of a sizable geographic area. The provision of this surveillance capability requires the evaluation of a large amount of information. In addition, to be of any practical use, the surveillance picture must not lag the actual air situation by any very appreciable interval of time. To accomplish this, the flight information must be gathered and evaluated very rapidly. This has been accomplished by minimizing the use of human beings in the surveillance system and leaning heavily on automatic methods. The various automatic data units in the SAGE System are, however, geographically scattered and must be interconnected so that they can process and exchange the flight information. This interconnection is made by special communication systems called data transmission systems which will be described herein.

Basic Considerations

There are several basic factors that have influenced the broad concept and ultimate design of these SAGE data transmission systems.[1] First and foremost among them is the fact that because of the very nature of the air surveillance problem, the SAGE System always works to some degree in real time. A second factor, so closely related to this as to be almost inseparable from it, is the accuracy with which the whole system must handle the continuous flow of information on which the air surveillance picture is based. A third factor of importance is the sheer size of SAGE. By any yardstick, it is a large project; geographically, for example, it covers the continental United States. Finally, and this is mainly of interest from the communication standpoint, the various data units to be interconnected are usually located in remote and sparsely settled areas.

Real Time

It is appropriate at this point to amplify the significance of the foregoing real-time factor somewhat, inasmuch as its impact on the design of a data transmission system may not be readily apparent. To those in the communications business, the real point of departure in data transmission system design is not only the type of information to be transmitted (analog or digital), but also whether or not the information is to be transmitted in real time. (The term "real time" is here loosely used.)

Real time is applied rather generally to those data transmission systems and data applications where the information being transmitted or handled has a high rate of obsolescence. An example of this is the information generated by a modern search radar tracking a particular aircraft. The position of the aircraft is known each time it is illuminated by the radar beam. This position report is obviously valid for only a short interval of time, and the higher the speed of the aircraft, the shorter this interval becomes. In fact, the position information is practically obsolete within the time of one radar sweep. Also the information has a very limited availability; it is generated only once during the radar sweep for any particular aircraft.

This is a fair sample of what is classed here as a real-time application. Most

---

Paper 58-332, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted November 25, 1957; made available for printing August 21, 1958.

R. T. JAMES is with the American Telephone and Telegraph Company, New York, N. Y.

In the preparation of this article, the author gratefully acknowledges the comments and suggestions of H. Sherman and R. Enticknap of the Lincoln Laboratories of the Massachusetts Institute of Technology, and of P. Mertz and D. Mitchell of the Bell Telephone Laboratories.

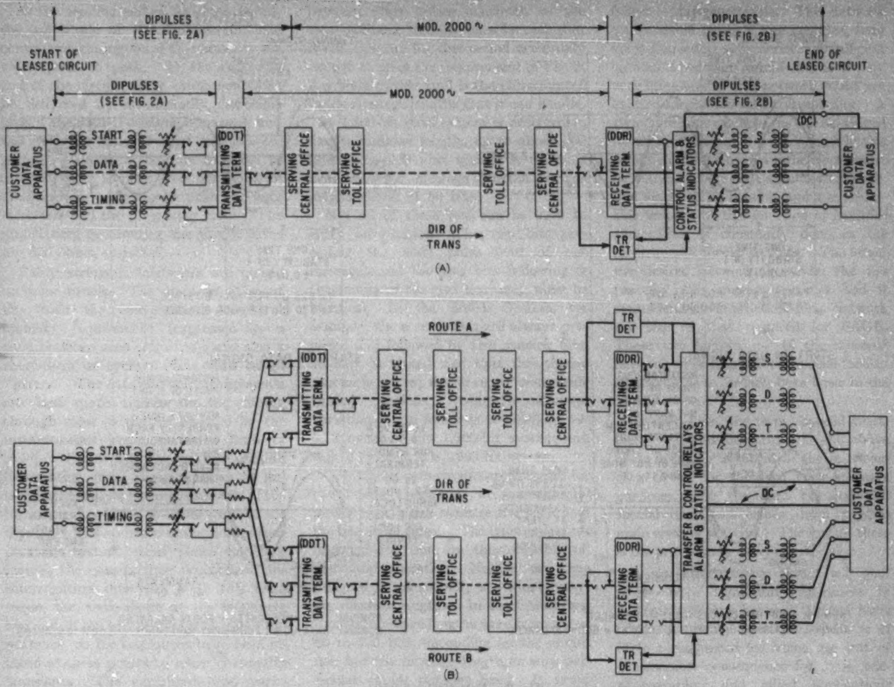

DIPULSES
(SEE FIG. 2A)
MOD. 2000 ∿
DIPULSES
(SEE FIG. 2B)
START OF
LEASED CIRCUIT
END OF
LEASED CIRCUIT
CUSTOMER DATA APPARATUS
START
DATA
TIMING
(DDT)
TRANSMITTING DATA TERM
SERVING CENTRAL OFFICE
SERVING TOLL OFFICE
SERVING TOLL OFFICE
SERVING CENTRAL OFFICE
(DDR)
RECEIVING DATA TERM
CONTROL ALARM & STATUS INDICATORS
(DC)
S
D
T
CUSTOMER DATA APPARATUS
TR DET
DIR OF TRANS
(A)
ROUTE A
TR DET
CUSTOMER DATA APPARATUS
START
DATA
TIMING
(DDT)
TRANSMITTING DATA TERM.
SERVING CENTRAL OFFICE
SERVING TOLL OFFICE
SERVING TOLL OFFICE
SERVING CENTRAL OFFICE
(DDR)
RECEIVING DATA TERM.
TRANSFER & CONTROL RELAYS ALARM & STATUS INDICATORS
S
D
T
DC
CUSTOMER DATA APPARATUS
DIR OF TRANS
(DDT)
TRANSMITTING DATA TERM.
(DDR)
RECEIVING DATA TERM.
TR DET
ROUTE B
(B)

Fig. 1. Data circuit arrangements. A—Single data circuit. B—Duplex data circuit

military data systems either for weapons control or surveillance are in this class; they vary only in the rate at which the information they use becomes obsolete and in the effects of obsolete or erroneous information on the operation of the overall system. It will be recognized that the real-time term as used here has various possible gradations, depending on the requirements of the particular system under consideration.

The main effect of this real-time factor on the design of a data transmission system is in the fundamental approach to the error problem. The data information is only available for transmission once, there is time enough to send it only once, and it can be received only once. In such a system, techniques of error correction tend to be of limited utility; techniques of error detection and eventual discard of erroneous information seem more practical and are of correspondingly greater importance. The incidence of errors throughout the system must, therefore, be kept small. In the interconnecting data systems this leads to the most rugged transmission systems possible consistent with bandwidth considerations, and to the control of noise in the transmission paths to the fullest extent practicable.

Error Rate

At present, excluding other connecting weapons systems, there are three main types of data systems used in the SAGE System. Grouped according to use, these are ground-to-ground, ground-to-air, and radar data systems. All of these systems do not work in real time to the same degree; neither do they have the same accuracy or error rate requirements. There may be actually two error rates to be considered for any particular system. The first is more or less conventional in that it is intended to be a short-term or minute-to-minute requirement: no more than one received error per 100,000 bits of information transmitted. The second error rate requirement is a longer term consideration and takes the form of a reliability requirement: no system outage or gross system malfunction to exceed 30 seconds.

The first of these requirements has been translated in terms of allowable circuit noise and has greatly influenced the engineering SAGE data systems. This was an early evalution of the short-term accuracy requirements; it is now generally appreciated that a blanket error rate of 1/100,000 represents a somewhat oversimplified view but it is still reflected in the engineering of all data systems. The second requirement is not applied to all data systems but only to those that are very important in the operation of the SAGE System.

Data Circuit Arrangements

Fig. 1 shows the two general data circuit arrangements used. (The data cir-

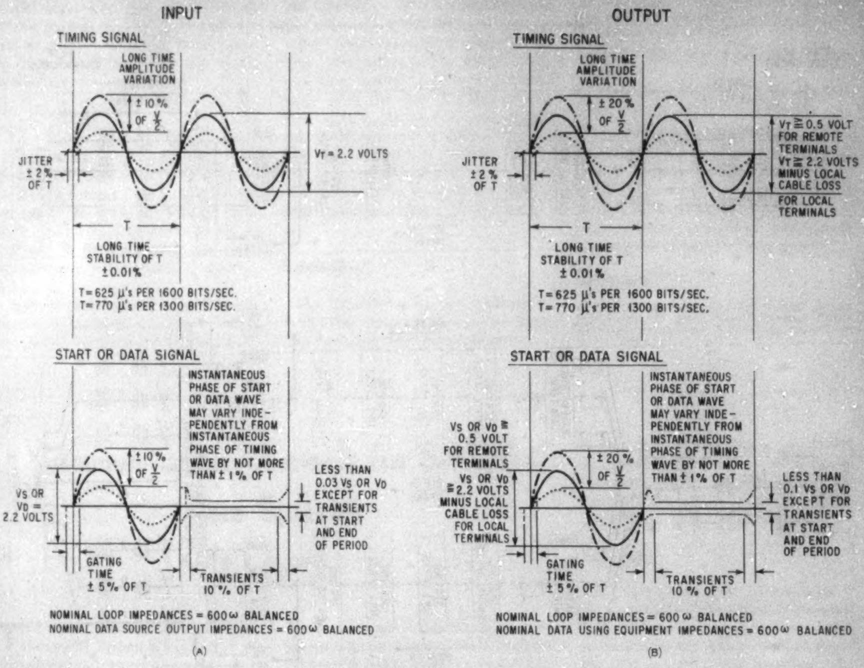

INPUT
TIMING SIGNAL
LONG TIME AMPLITUDE VARIATION
± 10 % OF V/2
JITTER ± 2% OF T
VT = 2.2 VOLTS
T
LONG TIME STABILITY OF T ±0.01%
T=625 μ's PER 1600 BITS/SEC.
T=770 μ's PER 1300 BITS/SEC.
START OR DATA SIGNAL
INSTANTANEOUS PHASE OF START OR DATA WAVE MAY VARY INDEPENDENTLY FROM INSTANTANEOUS PHASE OF TIMING WAVE BY NOT MORE THAN ± 1% OF T
± 10 % OF V/2
Vs OR VD = 2.2 VOLTS
LESS THAN 0.03 Vs OR VD EXCEPT FOR TRANSIENTS AT START AND END OF PERIOD
GATING TIME ± 5% OF T
TRANSIENTS 10 % OF T
NOMINAL LOOP IMPEDANCES = 600ω BALANCED
NOMINAL DATA SOURCE OUTPUT IMPEDANCES = 600ω BALANCED
(A)
OUTPUT
TIMING SIGNAL
LONG TIME AMPLITUDE VARIATION
± 20 % OF V/2
JITTER ± 2% OF T
VT ≧ 0.5 VOLT FOR REMOTE TERMINALS
VT ≧ 2.2 VOLTS MINUS LOCAL CABLE LOSS FOR LOCAL TERMINALS
T
LONG TIME STABILITY OF T ± 0.01%
T=625 μ's PER 1600 BITS/SEC.
T=770 μ's PER 1300 BITS/SEC.
START OR DATA SIGNAL
Vs OR VD ≧ 0.5 VOLT FOR REMOTE TERMINALS
Vs OR VD ≧ 2.2 VOLTS MINUS LOCAL CABLE LOSS FOR LOCAL TERMINALS
± 20 % OF V/2
INSTANTANEOUS PHASE OF START OR DATA WAVE MAY VARY INDEPENDENTLY FROM INSTANTANEOUS PHASE OF TIMING WAVE BY NOT MORE THAN ± 1% OF T
LESS THAN 0.1 Vs OR VD EXCEPT FOR TRANSIENTS AT START AND END OF PERIOD
GATING TIME ± 5% OF T
TRANSIENTS 10 % OF T
NOMINAL LOOP IMPEDANCES = 600ω BALANCED
NOMINAL DATA USING EQUIPMENT IMPEDANCES = 600ω BALANCED
(B)

Fig. 2. Data system signals

cuit terminal apparatus in each case is known as the "Western Electric *A1* digital data-signaling system.")[2] Circuit arrangement in Fig. 1(A) is used when the 30-second outage requirement is not applicable and normal circuit restoration times are satisfactory. Only one data path is provided; this arrangement is called a "nondual (or single) data circuit." In cases where the 30-second outage limit is applicable, the circuit arrangement shown in Fig. 1(B) is used. This is a dual (or duplex) data circuit. Two separate data transmission paths are provided between the data-processing units. At the transmitting end, both paths are fed simultaneously through a simple resistive network. At the receiving end of the circuit, trouble-sensing circuits (TR DET) capable of detecting major circuit troubles control a line switching arrangement that switches the data-processing apparatus to the alternate line whenever the working line goes in trouble.

A signaling system is also associated with this line switch circuitry that indicates the status of both transmission paths. Line transfers, although they result in a considerable number of errors, are accomplished in well under the 30-second limit. Substantially the same trouble-sensing apparatus is supplied with the nondual data circuits, excepting, of course, the line switch. In this application, status indications only are available. The actual physical routes used by both types of SAGE data circuits are chosen in accordance with certain military considerations. In the case of dual data circuits, the *A* and *B* routes are widely separate over all but a small part of the total distance between the data-processing units.

Data Circuit Characteristics

At the terminal points of the data circuits, information is exchanged with the data-processing units in the form shown in Fig. 2(A), the input, and (B) the output. Three separate informational components exist at these points: start, data, and timing. The start component indicates the beginning of a message, the data component contains the message information, and the timing component provides means of accurately synchronizing the data-processing units connected to the circuit.

Data circuits utilized in SAGE are designed to transmit information at two speeds, either 1,300 or 1,600 bits per second. The timing component is a continuous sine wave of either 1,300 or 1,600 cps (cycles per second) depending on the bit rate. The message information is carried by a series of marks and spaces on the data path. Marks are indicated by single cycles of either 1,300 or 1,600 cps frequency (dipulses), in the correct time slot. Spaces are no signal. The start of a message is indicated by a single mark on the start path, again a single cycle of either of the aforementioned frequencies depending on the bit rate. Fig. 2(A) and 2(B) illustrates a mark-space sequence on either of these last two paths.

At the sending end of the data circuit, the amplitude of the dipulses and the continuous timing wave is nominally 2.2 volts, peak to peak. At the receiving end of the circuit, these components can be delivered at substantially the same level if the data-processing apparatus and the receiving data terminal are close together. Otherwise they are adjusted to 0.5-volt, peak to peak, thus providing an allowance for path loss between the data terminals and the processing units. The impedances terminating the dipulse paths are 600 ohms, resistive.

Fairly stringent tolerances are placed on these signals. The timing component has both the conventional long-term stability requirement (expressed as a time between zero crossings), and also a short-term or cycle-to-cycle limit called "jitter." The data and start components are tied quite tightly to the timing through most of the bit interval by the instantaneous phase variation limit of ±1%. The data-processing units obtain information from these dipulses by interrogating only a few microseconds of the bit interval, usually at either peak of the dipulse. The accurate timing component plus the instantaneous phase tolerance insures the correct time position of this interrogation interval. With this technique, the wave-shape at the beginning and end of the bit interval is of minor importance, so the tolerances have been relaxed at these points to allow reasonable transients. The permitted level variations come out to be about ±1 db (decibel) at the data circuit input and ±1.5 db at its output. The allowable peak signal, peak noise ratios at these same two points, are 30 and 20 db respectively.

It may be noted that the differences between the dipulse signals at the input of the data circuit and those at its output are minor; the transmission is substantially distortionless. The dipulses, however, are transmitted as such over only a small part of the total length of the data circuit, at most a few miles. In the transmitting data terminal, the dipulse information is modulated on a suitable carrier (2,000 cps). This modulated signal is transmitted over specially treated voice bandwidth facilities to the receiving data terminal, where the information is extracted and regenerated in the original dipulse form.

This regeneration technique is also applied to the timing component which is regenerated by a local oscillator in the receiving data terminal that is periodically synchronized on each start pulse. The method of deriving timing limits the time between start pulses inasmuch as the local oscillator stability must be such that it can free-run for this period accurately enough to meet the requirement of Fig. 2. A system of this kind is therefore limited in the message lengths that it can handle. The 1,300-bit data system is designed to handle message lengths up to about 300 bits. The 1,600-bit system is designed primarily for message lengths in the neighborhood of 60 bits.

Not all of these bits can be used to carry information. The two bits preceding the start pulse (end of one message) and the two bits following it (beginning of the next message), must be identical. In the SAGE System, for example, the start pulses are always preceded and followed by two spacing bits. It may be noted also that this timing scheme is not capable of reproducing wide timing variations. About the maximum variation it can follow is ±0.01%, or ±0.13 cycle in the 1,300-bit system and ±0.16 cycle in the 1,600-bit system.

The SAGE data circuits are designed for continuous operation, i.e., meaningful messages or blank message frames are on the line at all times. The circuits are not designed for, nor are they capable of, start-stop operation. Neither are they designed to take rapidly and widely varying message lengths. In the SAGE System the message lengths vary from about 60 to 300 bits, depending on the circuit use, but the message length on any particular circuit remains fixed. It seems probable, however, that a given circuit could handle a moderate variation in message length although no performance tests have been made under these conditions. It may be noted here also that the same data terminal apparatus is not used for both the bit rates that have been discussed here. Different versions of the *A1* digital data signaling system are used for each speed.

Data Circuit Line Facilities

It is mentioned earlier that the sending and receiving data gear at the circuit terminals are interconnected by specially treated voice bandwidth facilities. It is now appropriate to describe briefly what this special treatment amounts to and how it affects the engineering of these line facilities.

Because of the extent of the SAGE System and the number of data circuits that it requires, it has seemed imperative from the beginning to avoid special treatment of the line facilities as far as possible. It has seemed equally imperative that the existing national network be used to the fullest extent practicable. This network comprises all types of line facilities, both voice frequency and carrier. Significant economic penalties would accrue if any very large slice of this network could not be used for SAGE data transmission. A given route, for example, may be developed for a certain type of carrier, and if this carrier cannot be used, another route must be found. To provide a limited network, this may well be possible and a considerable amount of this selection of facilities and routes is constantly done in the furnishing of special services. The larger the desired network, however, the less practical this process becomes and it seems entirely impractical for a network the size of that required for SAGE. These two factors, use of the national network and a minimum of special facility treatment have, in fact, been basic in the SAGE concept.

As used here, the term "special treatment" includes measures over and above those normally involved in the provision of a circuit for standard voice message purposes. For the data circuits, this special treatment comes down to two main areas of activity. The first of these is the provision of delay equalization.

A standard message circuit will generally have significant differences in transmission time between various parts of the frequency spectrum. This is of little consequence for voice use but of considerable consequence for data, telephotography, and allied applications where the rate of information flow is much higher. This difference in transmission time, or delay, varies with the type of facility in question, and in some types of facilities it also varies with the circuit length. This distortion results in different frequency components of the signal arriving at the receiving end of the circuit at different times, and unless it is corrected, a degradation of the signal results.

To alleviate the effects of this distortion, it is customary to use only the part of the available circuit bandwidth that can be readily equalized. The least distortion will be found in the middle range of the circuit bandwidth; it is much greater in the upper and lower circuit cutoff regions. In the SAGE data circuits, these cutoff regions are not used at all, the line signal energy of the *A1* data system being confined to the region from 1,000 to 2,600 cps. (The signal pass band must be limited and shaped, in any case, for other reasons.) The delay distortion in this region is held to within ±250 microseconds; in most cases, delay equalization must be added to keep within this limit. The limit itself was set on the basis of telephotograph experience and theoretical consideration of the SAGE requirements.

Each type of line facility has an individual and distinctive delay distortion characteristic. It is possible, therefore, to predict the distortion characteristic for any data circuit providing its make-up (the types and lengths of line facilities comprising it), is known. With the distortion characteristic known, it is entirely practicable to apply the appropriate equalization to reduce it. Circuits can be equalized in this manner to a reasonable degree of precision if the circuit make-up information is accurate and if no impedance irregularities exist. This is the general method used to equalize the data circuits for SAGE; it is a prescription process rather than a measurement process. The technique also includes procedures designed to expose any impedance irregularities that may exist; this has been found to be essential. Not all circuits can be equalized with this method, some makeups are so complex that an actual measurement must be made. Fortunately, this holds for only a small minority of the data circuits. In conclusion, it may be noted that the delay distortion characteristics of the circuits are fairly stable and, for the circuit lengths involved in SAGE, relatively little seasonal change may be expected.

The second major area of effort has been in the control of circuit noise. As pointed out earlier, SAGE data circuits work to some degree in real time and the emphasis has necessarily been on cutting down the incidence of errors rather than their correction. This, in the end, becomes a problem of circuit noise.

It may be noted at the outset that the national network was, and is, still being built primarily for the transmission of information via the spoken word. Any added use of it, such as data transmission, involves an adaptation process to reconcile conflicting technical requirements or to meet new ones. Although this may be a troublesome job, the economic advantages attendent on it make it well worth while.

The noise situation is a good example of this. The frequency response of the human ear is not flat but is peaked in the middle range of audible frequencies, around 1,000 cps. Above and below this, its response falls off, particularly so in the region below about 500 cps and above about 2,500 cps. This has been considered in the engineering and construction of the national network and its basic noise characteristics reflect this kind of a weighting. The net result is that any other user of the network that does not have this built-in-weighting is likely to encounter comparatively high noise, particularly at the low end of the spectrum. Another human characteristic of interest in this regard is the response time or "time constant" of the ear. It has been found that noise meters with a time constant of 200 milliseconds (plus the appropriate weighting) give the most accurate gage of the effect of circuit noise on normal voice transmission. Noise bursts of shorter duration than 200 milliseconds have correspondingly less effect and their effect continues to decrease as their duration time shortens. This, also, has been reflected in the engineering of the national network so that, in addition to their weighting, the circuit noise characteristic may show sporadic short-noise peaks that may range anywhere from 10 to 50 db above the average or steady-noise level.

The rate of occurrence of these noise bursts, called impulse type noise, as well as their peak factor, will vary with the type of facility. Actually, it depends on the noise exposure; some types of facilities, particularly some types of carriers, work under more severe noise exposures than others because of their area of application in the telephone plant. The sources of this impulse noise are many and varied, the two main causes being atmospheric static disturbances and circuit switching operations.

So far as the SAGE data systems are concerned, the effect of both the network noise weighting and the average circuit noise level is not controlling. As noted, the signal spectrum is carefully shaped and limited, so the *A1* system normally operates in the part of the circuit bandwidth where the noise is most carefully controlled. Furthermore, the weighted average noise level on any normal circuit is not likely to be such that it will impair the operation of the data system. The level of this noise is held to that acceptable to human beings and they are much less tolerant than the SAGE data systems in this respect.

The opposite is true of impulse noise. The SAGE data systems are much less tolerant of this type of noise than human beings. The information carried by these systems is in digital form with the bit intervals less than one millisecond in length. The level of impulse noise of this or shorter duration intervals accordingly constitutes our main concern in the whole noise area. It has recently been determined that errors in transmission generally occur only when the magnitude of individual noise peaks exceeds about −15 dbm (decibels referred to one milliwatt) at the zero dbm transmission level point. Plans now under preparation make necessary modifications in various techniques of circuit design and noise measurement to reflect this criterion. The design techniques vary with the facility involved; most of the work in this area concerns the shorter haul carrier systems because they operate at lower line levels than have heretofore been employed.

The accomplishments in both the noise and delay problem areas have borne fruit in making possible the widespread use of the company's national network for SAGE data circuits. All the standard Bell System type of facilities except circuits with heavier loading than *H* or *B88*, and one older type of carrier, are being used for this purpose at present. These exceptions do not represent a significant part of the network.

One further point may be noted with respect to facilities. The choice of dipulse signals as the input and output signals of the data circuits eases considerably the problem of connecting the data circuits and the data-processing apparatus. In most cases, separation of the data circuit terminals and the data-processing apparatus has been found to be desirable and necessary. These dipulse signals permit this; they can be transmitted quite readily for several miles if necessary.

Performance

This summary of the main features of the data circuits provided for the SAGE System would not be complete without a few words as to their performance. Only a limited number of circuits are presently in service but these results are also being supplemented by a rather extensive field testing program designed to check error performance. On the basis of this experience, it seems that the 1/100,000 blanket error rate, roughly one error per minute, is generally met if the errors are averaged over a fairly long period, at least several hours. The normal circuit performance is not a fairly constant incidence of errors from minute to minute but many minutes of absolutely error free transmission followed by a burst of errors, or errors equivalent to a burst, then error-free transmission again.

Conclusions

Transmitting information in real time seems likely to be a problem for some time to come. The SAGE System is only one example of it; many more applications presently exist or are in the planning stage. This paper represents one approach to the problem; other approaches are possible and are, in fact, being tried. It seems probable that the experience gained in this and other data systems will ultimately indicate the best technical and economic solution of the real-time data problem.

References

1. SAGE Data System Considerations, R. G. Enticknap, E. F. Schuster. *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 824-32.

2. SAGE Data Terminals, R. O. Soffel, E. G. Spack. *Ibid.*, pp. 872-79.

The Ferroresonant Circuit

GEORGE E. KELLY, JR.
ASSOCIATE MEMBER AIEE

A CIRCUIT which consists of capacitance in series with an iron-core reactor, with or without other circuit elements, may be a ferroresonant circuit. Ferroresonance, as the term indicates, is resonance associated with circuits in which at least one of the circuit elements contains iron and which is a nonlinear inductive element. Resonance in any series circuit occurs where the inductive and capacitive reactances are equal. In a given series circuit of linear resistance, inductance, and capacitance, there is only one frequency of applied sinusoidal voltage at which resonance will occur; but at this frequency, resonance occurs at all values of applied sinusoidal voltage.

A circuit with linear resistance and capacitance in series with an iron-core reactor, regardless of the frequency of the applied sinusoidal voltage, will be a ferroresonant circuit if, at some value of applied voltage, the inductive and capacitive reactances are equal. The resistance and reactance of the iron-core reactor vary with the degree of saturation in the iron, which depends upon the voltage across the reactor. At negligible saturation, the inductive reactance is high but it decreases as saturation increases. When the capacitive reactance is lower than the inductive the resulting current lags the applied voltage, but if it is higher the current will lead the applied voltage. Harmonics will be present because of the nonlinear inductance but, as will be explained later, their presence does not greatly affect the performance of a ferroresonant circuit except in the region of high saturation.

Fig. 1(A) shows sustained rms current of substantially fundamental frequency plotted versus applied sinusoidal voltage in a ferroresonant circuit. At point y on the ys curve, corresponding to voltage B, the voltage and current are substantially in phase. This is the point of ferroresonance. For other points on the ys curve, the current leads the applied voltage. For values of applied voltage between B and C, there are two possible values of substantially fundamental frequency current, either of which may occur depending upon the residual magnetism in the reactor and the charge on the capacitor at the instant the voltage is applied.

Assume that a value of voltage less than voltage B of Fig. 1(A) is applied to a ferroresonant circuit, and subsequently the applied voltage is increased; a curve similar to $0x$ is traced. At point x, I suddenly jumps to curve ys with the same value of applied voltage corresponding to I at point x on the $0x$ curve. If the circuit is kept closed and the applied voltage varied, curve ys is traced. Under the preceding condition, the circuit will operate to return to curve $0x$ if the applied voltage is decreased to point y on the ys curve, corresponding to applied voltage B. The return from curve ys to $0x$ may also be accomplished if the circuit is opened and closed one or more times in the region between applied voltages corresponding to points B and C.

Fig. 1(A) and the foregoing description of the simplest type of the ferroresonant circuit may be obtained from previously published papers,[1–9] where the possibility of subharmonic currents and voltages is also discussed.

Previous Investigation

Several of the many papers just mentioned dealing with the ferroresonant circuit have been delivered before the AIEE.[1–9] As indicated by their titles, they are chiefly concerned with subharmonic voltages and currents which may be present. The usual subharmonic components in a 60-cycle circuit are those of the second and third order, with frequencies of 30 and 20 cycles per second respectively.

The ferroresonant circuit is of interest because of its possible application to relays and control devices,[1,3,4] and because of its application, when extended, to the problem of open conductors in 3-phase circuits. In this latter application, interest is centered on the magnitude of the resultant voltage across the nonlinear reactance, and its phase relative to that of the impressed voltage.

McCrumm,[9] in reporting the results on his extensive tests on ferroresonant circuits, although chiefly concerned with subharmonics, gives a clear picture of what may be expected in a ferroresonant circuit. He gives a graph of impressed sinusoidal voltage E versus rms current I in the circuit. A similar graph is shown in Fig. 1(A).

As the impressed sinusoidal voltage is slowly increased from zero to its maximum value, the response of the circuit (its rms current and rms voltage across the reactor) differ in character for different magnitudes of impressed sinusoidal voltage. Referring to Fig. 1(A), it may be seen that for any value of E between 0 and B there can be but one

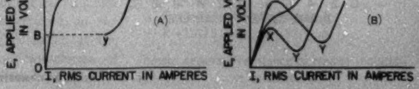


Fig. 1. Current-voltage characteristics of a ferroresonant circuit

---

Paper 58-1168, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted January 22, 1957; made available for printing July 30, 1958.

George E. Kelly, Jr. is with the Prairie View Agricultural and Mechanical College, Prairie View, Tex.

The author gratefully acknowledges the helpful suggestions of Profs. Edith Clarke and William C. Duesterhoeft of the University of Texas, in the preparation of this paper.

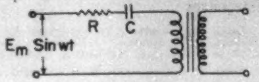


Fig. 2. A ferroresonant circuit

value of *I* (a possible exception will be mentioned later). For any value of *E* above *C*, there is but one possible value of *I*. For values of *E* between 0 and *B*, *I* lags *E*; for values of *E* above *C*, *I* leads *E*. For values of *E* between *B* and *C*, Fig. 1(A) shows two possible values of *I*. In addition to these two values, McCrumm has shown that subharmonics of both the second and third order may occur for values of *E* between *B* and *C*, and also for values of *E* below *B*.

In a paper published prior to McCrumm's, Odessey and Weber[5] give methods for determining critical values, or points of discontinuity. The points *x* and *y* on the curve 0*x* and *yz* of Fig. 1(A) are points of discontinuity. They also show a graph of curves similar to the curve of Fig. 1(B), and point out that the sections of the curves where the rate of change of current with applied voltage is negative, although satisfying voltage-current equations, does not give stable operating points. In Fig. 1(B), points on the curve between *x* and *y* are unstable points. This agrees with the graph of Fig. 1(A) where no connection is shown between points *x* and *y*.

In some of the published papers on the ferroresonant circuit, extensive test data are given, without accompanying calculations to check test data; in others, equations are developed based on theory without accompanying tests. So far as the author has been able to determine, no publication gives both test results and calculations to check test results for a specified ferroresonant circuit. Because of this lack, the present study was undertaken.

Purpose of This Study

This study of the ferroresonant circuit was undertaken for the following purposes:

1. To determine whether the rms currents in the circuit and the rms voltages across the nonlinear reactor, under various values of applied sinusoidal voltage, can be satisfactorily predicted from calculations in which rms currents and voltages are treated as if they were of fundamental frequency only.
2. To predict the probability of the occurrence of each of two possible voltages across the reactor in a region where either is possible.
3. To observe subharmonic voltages across the nonlinear reactor, if present, and to record the conditions under which they occur.
4. To illustrate the application of the ferroresonant circuit to the problem of open conductors in 3-phase circuits.

The paper is divided for convenience into four parts, each dealing with one of the stated purposes for which this study of the ferroresonant circuit was made.

As the same type of circuit and the same iron-core reactor were used throughout, they will be described before stating the methods used in calculations and in making tests.

The circuit studied consisted of resistance and capacitance in series with an iron-core reactor, the iron-core reactor being the 120-volt winding of a 2-kva, 120- to 240-volt, single-phase transformer with the 240-volt winding on open circuit. A 60-cycle sinusoidal voltage of variable magnitude was applied to this circuit, in which resistance and capacitance were given several different values; see Fig. 2.

The excitation characteristics of the reactor are given in Fig. 3. In Fig. 3(A), rms exciting current is plotted versus rms applied sinusoidal voltage; in Fig. 3(B) and (C) respectively, rms magnetizing and power components of exciting current are plotted versus the applied rms sinusoidal voltage.

Calculations and Test for the Same Ferroresonant Circuit

Two methods of calculation were used, both making use of the excitation characteristics of the reactor. In both methods, all currents and voltages were treated as if they were of fundamental frequency only.

In the first method it was assumed that the total exciting current in the reactor could be considered magnetizing current, with the reactor voltage leading it by 90 degrees. For any assumed reactor voltage, with the corresponding exciting current as reference vector, the vector voltage across the resistance and 60-cycle capacitive reactance in series was added vectorially to the reactor voltage. This gave the applied voltage *E* necessary to produce the assumed reactor voltage. In Fig. 4(A), curve *A* gives the magnitude of the voltage *E*, calculated by the first method.

In the second method, a reactor voltage was assumed and the corresponding magnetizing and power components of the rms exciting current were used to determine the phase of the exciting current relative to the reactor voltage. With reactor voltage as reference vector, the voltage across the resistance and 60-cycle capacitive reactance in series caused by the vector exciting current was added to the reactor voltage to give the applied voltage *E* required to produce the assumed reactor voltage. The phase of the reactor voltage relative to *E* was

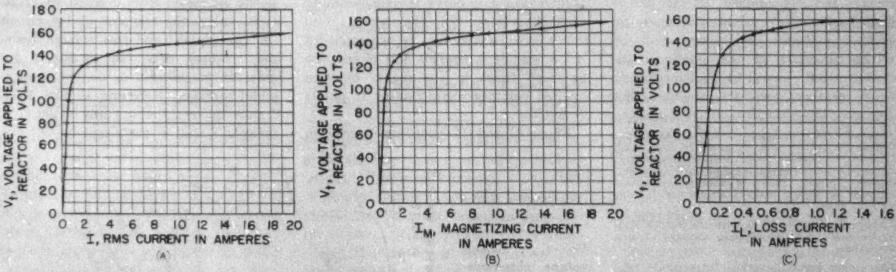
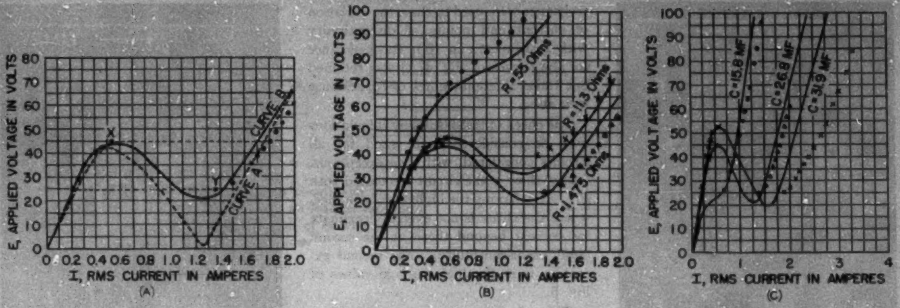


Fig. 3. Excitation characteristics of an iron-core reactor



(A)—Curve A: Apparent resistance of reactor neglected
Curve B: Apparent resistance of reactor included

o—test values
C=26.8 μf (microfarads)
R=1.475 ohms

(B)—C=26.8 μf

Note: apparent resistance of reactor included in calculations

(C)—R=1.475 ohms

Note: apparent resistance of reactor included in calculations

Fig. 4. Current-voltage characteristics of a ferroresonant circuit

also determined. In Fig. 4(A), curve *B* gives the magnitude of the rms current *I* versus the magnitude of the voltage *E*, calculated by the second method.

First Method of Testing

Two methods of testing were employed. The first was used to determine whether test results can be checked, at least approximately, by calculations based on the assumption that rms currents and voltages are of fundamental frequency only. By this method, the procedure was as follows: The switch to the circuit was closed and kept closed, while the applied sinusoidal voltage was varied from zero to a high value. Applied voltages and resultant rms currents were recorded. Then, with the switch to the circuit still closed, the applied voltage was gradually decreased. Applied voltages and rms currents were again recorded.

Results of First Method of Testing

The results of one test by the first method is shown in Fig. 4(A), where rms current *I* is plotted versus applied rms sinusoidal voltage *E*. As *E* was slowly increased up to approximately 45 volts, a smooth curve of *I* versus *E* was obtained; but at $E=45$ volts, *I* suddenly jumped to a higher value. After that, as *E* was further increased, another section of a smooth curve was obtained, as indicated by the small circles. As the voltage *E* was then slowly decreased, *I* retraced its previous path to the point where *E* was approximately 45 volts. It did not jump to a lower value at this point, but continued on the same smooth curve until *E* was approximately 25 volts, and *I* approximately 1.36 amperes. At this point it suddenly jumped to 0.21 ampere at 25 volts. From this point, it retraced its previous path down to zero.

Figs. 4(B) and (C) show the results of additional tests and calculations, with values of resistance and capacitance indicated.

Two Possible Values of Current and Reactor Voltage for One Value of Applied Voltage

As may be seen from Fig. 4(A), if the applied voltage *E* is between 25 and 45 volts, there are two possible values of current. There are also two possible values of voltage across the reactor. This is shown by curve *A* of Fig. 5, where rms voltage across the reactor is plotted versus applied sinusoidal voltage *E*. Figs. 6(A) and 6(B), reading time from left to right, show one period of each of the two possible reactor voltage waves superposed upon one period of the applied sinusoidal voltage of 45 rms volts. The voltage at low current is given in (A); that at high current, in (B). Although the time phase of each reactor voltage wave relative to the applied voltage is correct, no attempt was made to have their amplitudes to the same scale.

In Fig. 7(A), taken at the high current value, the waveforms of the current and reactor voltage are superposed on the applied voltage. In (B), taken at the low current value, the waveforms of the current and reactor voltage are superposed on the applied voltage wave. These curves were also taken at the point of discontinuity, corresponding to an applied voltage of 45 rms volts, and the circuit associated with Fig. 4(A).

Comparison of Calculated and Test Values with Conclusions

For the circuit associated with Fig. 4(A), both methods of calculation, previously described, were used. For applied voltages between 0 and 45 volts,

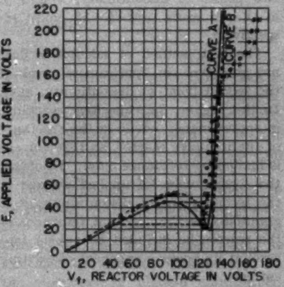


Fig. 5. Voltage characteristics of a ferroresonant circuit

Calculated curves:
A—R=1.475 ohms, C=26.8 μf
B—R=1.475 ohms, C=31.9 μf

Test values:
o—R=1.475 ohms, C=26.8 μf
x—R=1.475 ohms, C=31.9 μf

Note: apparent resistance of reactor included in calculations

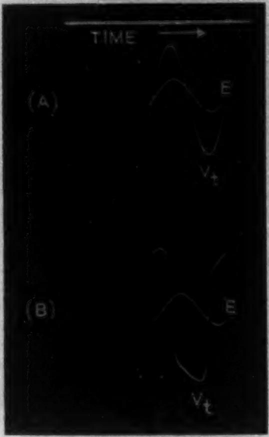


Fig. 6. Oscillograms illustrating phase relationship between applied and reactor voltage before and after jump from low to high current of ferroresonance at point of discontinuity. R=1.475 ohms; C=26.8 µf

the second method of calculation (where the power component of the reactor exciting current is taken into account) gives a better check of test results than the first method (where the exciting current is treated as entirely magnetizing current). It also gives a better check of the two points of discontinuity, x and y, at $E=45$ volts and $E=25$ volts respectively.

The first method of calculation does not give even an approximate check of the second point of discontinuity, at y. In this region, the capacitive reactance is approximately equal to the reactance of the iron-core reactor, so that current is limited by resistance only. To obtain the correct value of applied voltage at the second point of discontinuity, it is essential that the apparent resistance of the reactor be included by taking into account its power component of exciting current.

Because the first method of calculation is easier to apply, it was used in the calculations for Fig. 4(A), the first study made. It is of interest to note that for high values of current it gives a better check than the second method.

In Fig. 4(B) and (C), the second method of calculation only was used, and test results for each case are indicated. In Fig. 4(B), the capacitance used for the curves in Fig. 4(A) was retained, but three different values of resistance were used to show the variation of current and points of discontinuity with resistance. In (C), the resistance for the curves of Fig. 4(A) was retained but three different values of capacitance were used.

The check of test results by calculations may be considered quite good for all circuits considered, with two exceptions. One of these is the circuit associated with Fig. 4(C), having the highest capacitance, where the check is not even approximate for current beyond those at the second point of discontinuity. The second exception is in the high voltage values of Fig. 5.

This second exception is not surprising. At the first point of discontinuity the voltages across the reactor, as shown in Fig. 6(A) and (B), before and after the discontinuity respectively, are approximately sinusoidal and of fundamental frequency. Therefore the method of calculation is reasonable. As the applied voltage is increased, the waveshape does not change appreciably until the applied voltage is far above the rating of the transformer. Then it changes suddenly and, although of the same period, has pronounced harmonics, including even ones. If it is remembered that the applied voltage is from a power source at constant voltage, the application of voltages high above rated transformer voltage is not a practical case in power transmission systems.

A more thorough study should be made of these two exceptions, both of which involve high exciting currents in the reactor, with the idea of developing a method of calculation which will include the effect of harmonics when exciting currents are above a predetermined value.

The phase positions of the reactor voltage waves relative to the applied voltage, shown in Fig. 6(A) and (B) at an applied voltage of 45 volts, agree with calculations. For low values of current, the reactor voltage is almost in phase with the applied voltage. For high values of current, it is almost 180 degrees out of phase with the applied voltage. Thus, there is phase reversal at the points of discontinuity.

A more explicit explanation of linear calculations giving a good check of experimental results is as follows. It may be recalled that a nonlinear reactor generates harmonics, these harmonics being most pronounced in the region of saturation. In view of the preceding statement, it would be expected that linear calculations would give a good check of experimental results, provided the region of high saturation or pronounced harmonics is not entered. Such was the case in this investigation.

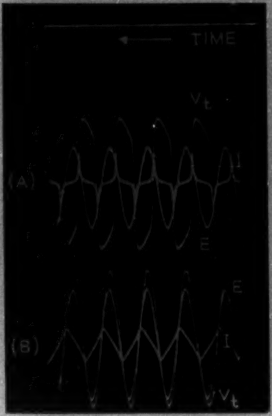


Fig. 7. Same as Fig. 6, illustrating phase relationship between applied voltage, reactor voltage, and current

A—$E=45$ volts
$V_t=122$ volts
$I_{rms}=1.80$ amperes
B—$E=45$ volts
$V_t=87$ volts
$I_{rms}=0.53$ ampere

Probability of Occurrence

The probability of occurrence of each of the two possible voltages across the reactor when either is possible was tested. The circuit associated with Fig. 4(A) was employed for this test. In this figure it may be seen that for values of applied voltage between 25 and 45 volts there are two possible values of current. There are also two values of voltage across the reactor, as shown in Fig. 5. The question is which is more likely to occur. To answer this question, four types of tests were made.

TEST 1: SWITCH CLOSED AND VOLTAGE VARIED

As the applied voltage was slowly raised from zero to 45 volts, without opening the switch, the lower value of reactor voltage was obtained. At 45 volts there was a jump to higher current and higher voltage across the reactor, and phase reversal. With the switch still closed, and the applied voltage reduced, the higher values of current and

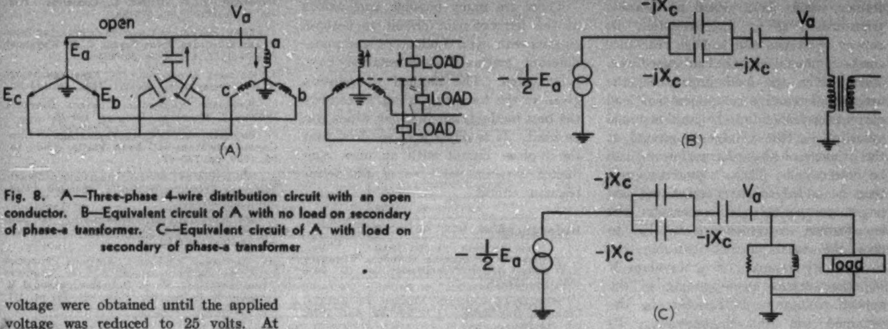


Fig. 8. A—Three-phase 4-wire distribution circuit with an open conductor. B—Equivalent circuit of A with no load on secondary of phase-a transformer. C—Equivalent circuit of A with load on secondary of phase-a transformer

voltage were obtained until the applied voltage was reduced to 25 volts. At this second point of discontinuity there was jump to lower values of current and voltage.

It may be concluded from this first test that, once on a higher or lower voltage path, with the switch closed, there will be no change to the other until points of discontinuity are reached.

TESTS 2 TO 4: VOLTAGES APPLIED TO CIRCUIT BY SUDDEN OPENING AND CLOSING OF SWITCH

Three different tests were made for applied voltages 35, 39, 41, and 43 volts rms, including the following initial conditions: 1. probable residual magnetism, but no charge on the capacitor; 2. no residual magnetism and no charge on the capacitor; and 3. probable residual magnetism and charge on the capacitor.

In tests with initial conditions 1 and 2, the switch was opened and closed 100 times. In the test with initial condition 3, it was thought that the voltage across the reactor corresponding to high or low currents of ferroresonance just before opening and then closing the switch would influence results. Therefore, a set of tests was made for each applied voltage. In the first, the voltage across the reactor corresponded to a point on the curve, a low current of ferroresonance; in the second, a high current of ferroresonance.

In these tests, no attempt was made to determine the initial value of the applied sinusoidal voltage. However, it is here concluded that for applied voltages between 25 and 35 volts, the low voltage values are more likely to occur than the high. For applied voltages between 41 and 45 volts, the high voltage values are more likely to occur than the low. The preceding conclusion is further justified in view of the fact that, inductance being a function of current, the closer saturation is approached the more predominating capacitive reactance becomes on the circuit impedance, thereby idealizing the conditions necessary to substantiate high current of ferroresonance.

Subharmonics

No sustained subharmonics were observed at any time during the tests. In the preceding section, where voltages were suddenly applied to the circuit, with and without discharging the capacitor, and with and without eliminating residual magnetism in the reactor, there was no indication that subharmonics were present, even as transients. This is not surprising, as it has been stated that subharmonics occur only at suitable values of switching angle, charge on the capacitor, and flux linkages for the necessary degree of saturation in the iron-core reactor.[9]

The circuit was closed hundreds of times in the tests described in the preceding section, but without regard to the initial value of the applied voltage. It would seem that on at least one occasion the initial value of the applied voltage would have been suitable for subharmonics. As none appeared, it can be concluded that their occurrence in the particular circuits studied would be very rare indeed, if they occurred at all. The states of ferroresonance and subharmonic oscillations are totally different phenomena unrelated in behavior.[9]

Three-Phase Circuit with Open Conductor as Example of Ferroresonant Circuit

In a recent paper,[10] tests and calculations to check tests are given for a 3-phase circuit with open conductors. The circuit considered is a 3-phase 4-wire distribution circuit, shown by a 3-line diagram in Fig. 8(A). Balanced sinusoidal voltages to ground E_a, E_b, and E_c are applied to the grounded Y-connected primary of a transformer bank, paralled by a Y-connected capacitor bank, with grounded secondary supplying a 3-phase 4-wire distribution circuit.

Let phase *a* be opened between the power source and the capacitor bank, as in Fig. 8(A).

With phase *a* at the power source as reference vector, the applied voltages E_b and E_c are

$$E_b = -\frac{1}{2}E_a - j\frac{\sqrt{3}}{2}E_a$$

$$E_c = -\frac{1}{2}E_a + j\frac{\sqrt{3}}{2}E_a$$

The voltage

$$E_c - E_b = j\sqrt{3}\,E_a$$

sends no current to ground, nor does it send any current into the open phase *a*. It can therefore be ignored in determining the voltage to ground V_a at the primary terminal of the transformer in phase *a*.

The component $-1/2E_a$, common to E_b and E_c, will send current to ground through the capacitor bank and phase-*a* primary of the transformer bank, as indicated by arrows. With no load on the secondary of phase-*a* transformer, the circuit to be studied is the simple ferroresonant circuit shown in Fig. 8(B). With load on the secondary, it is the circuit of Fig. 8(C).

In both circuits, the capacitive reactance is $-j3/2x_c$. For the case of no load on phase-*a* transformer, the circuit is similar to those studied in the section "Calculations and Test for the Same Ferroresonant Circuit" there is no resistance but that of the transformer itself. With load on phase-a transformer, as an approximation, its copper resistance and leakage reactance can be combined with the load impedance. Then, given the load impedance, the active and reactive components of load current corresponding to any assumed value of V_a (the voltage to ground at the primary of phase-a transformer) can be determined. These components can then be added to the magnetizing and power components of current in the transformer, corresponding to V_a, to give the total current entering the circuit. By assuming several values of V_a, the current corresponding to the applied voltage $-1/2E_a$, can be determined, and from it, the voltage V_a at the transformer.

There are many possible applications of the ferroresonant circuit to 3-phase circuits with open conductors, and many different methods of calculation have been given. The method of solution given in the foregoing is not necessarily the best method, but it is one which can be used. It is given here to show that the 3-phase circuit with an open conductor is a special case of the ferroresonant circuit

References

1. Studies in Nonlinear Circuits, Chauncey Guy Suits. *AIEE Transactions*, vol. 50, June 1931, pp. 724–36.

2. Nonlinear Circuits Applied to Relays, Chauncey Guy Suits. *Ibid.*, vol. 52, Apr. 1933, pp. 244–46.

3. Analysis of Series Capacitor Application Problems, J. W. Butler, C. Concordia. *Ibid.*, vol. 56, Aug. 1937, pp. 975–88.

4. Subharmonics in Circuits Containing Iron-Cored Reactors, Irven Travis, C. N. Weygandt. *Ibid.*, vol. 57, Aug. 1938, pp. 423–31.

5. Critical Conditions in Ferroresonance, P. H. Odessey, Ernst Weber. *Ibid.*, pp. 444–52.

6. Resonant Nonlinear Control Circuits, William T. Thomson. *Ibid.*, pp. 469–76.

7. Subharmonics in Circuits Containing Iron-Cored Inductors—II, Irven Travis. *Ibid.*, vol. 58, 1939, pp. 735–42.

8. Similitude of Critical Conditions in Ferroresonant Circuits, William T. Thomson. *Ibid.*, Mar., pp. 127–30.

9. An Experimental Investigation of Subharmonic Currents, John D. McCrumm. *Ibid.*, vol. 60, 1941, pp. 533–40.

10. Abnormal Voltage Conditions Produced by Open Conductors on 3-Phase Circuits Using Shunt Capacitors, P. E. Hendrickson, I. B. Johnson, N. R. Schultz. *Ibid.*, vol. 72, pt. III, Dec. 1953, pp. 1183–93.

(See Discussion on Page 1061)

Recent Progress in Applying Information Theory to Digital Transmission Systems

R. FILIPOWSKY
ASSOCIATE MEMBER AIEE

Synopsis: Recent investigations, still being continued at Westinghouse, seem to indicate that careful observation of the detection theory, of the definition of digital systems, of the size of the signaling alphabet, and of the conversion process from low-pass signals to band-pass (carrier) signals may be extremely helpful for faster progress in applying information theory to digital communication systems.

MANY are convinced that information theory presented very useful results in the last decade and the author believes that many tried to apply these results in practical communication system design. The title of the paper is not meant to imply any announcement of spectacular new methods of information transmission or a presentation of final results of the author's current investigations. In this paper the author wishes to show that careful observation of all assumptions under which theoretical results have been derived does simplify and accelerate the process of applying information theory to digital communication systems. In particular the following points are stressed:

1. The detection problem for communication systems is practically in all aspects different from the detection problem in radar.

2. The term digital transmission requires an accurate definition and many different types of digital transmission systems are conceivable.

3. It seems that when using larger signal alphabets, in particular when selecting for each elementary interval in a digital system one and only one waveform out of a larger set of uniquely specified waveforms, one may arrive at more efficient digital systems, than when operating on a binary basis.

4. In truly digital systems one should specify directly carrier-band waveforms in place of a particular modulation method for carrier transmission.

With regard to item 1, the reader may remember that radar has led the way in applying the theory of optimum detection systems. Digital transmission systems have at least the pulsive operation in common with radar and it is therefore logical to expect that results of optimum radar detectors could be applied to data transmission systems.

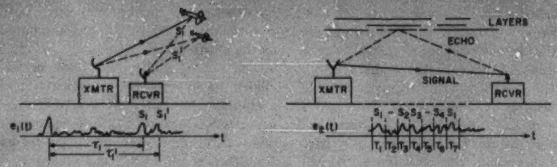
Fig. 1 is a comparison of the most important features of a typical radar system with those of a typical communication system. Notice that practically all design parameters are of opposite character in the two cases.

Receiver and transmitter are generally at the same location in radar, but are far apart in communication links. Information is carried by the energy radiated from a communication transmitter, whereas it is generated within the channel in radar systems. Only one exactly known typical signal is transmitted in radar, but selections out of many different types of such signals, not all completely predictable, may be emitted by a communications transmitter. The detection problem in radar is primarily restricted to optimum discrimination between this one a priori known waveform and Gaussian noise. In communications, it is much more complicated, involving the error-free discrimination between all possible types of signals or the distortion-free extraction of continuous signals from noise. The time of arrival of a signal is not known in radar, because this is actually the information-carrying characteristic, which should be extracted as accurately as possible. It may be completely known in advance in some communication systems and this knowledge may be used to improve the detectibility, i.e., by clocked operation or synchronous detection. The direction of arrival is unpredictable in radar and carries another part of the wanted information. It is generally known in communication systems, except in mobile systems, where it can only be predicted by co-operation between both stations. The signal-to-noise ratio is down to the lowest possible value in radar, where it determines the

---

Paper 58-1157, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted April 11, 1958; made available for printing July 25, 1958.

R. Filipowsky is with Westinghouse Electric Corporation, Baltimore, Md.



	CHARACTERISTIC	RADAR	COMMUNICATIONS
1.	LOCATION	XMTR CLOSE TO RECVR	BOTH FAR DISTANT
2.	INFORMATION	CREATED IN CHANNEL	RADIATED FROM XMTR
3.	TYPES OF SIGNAL	ONE OF KNOWN WAVESHAPE	MANY, NOT ALL PREDICTABLE
4.	DISCRIMINATION	PRIMARILY SIGNAL VS. NOISE	SIGNAL VS. SIGNAL AND VS. NOISE
5.	TIME OF ARRIVAL	NOT KNOWN, CARRIES INFORMATION	KNOWN IN "CLOCKED" OPERATION
6.	DIRECTION OF ARRIVAL	VARIABLE, CARRIES INFORMATION	GENERALLY KNOWN AND FIXED
7.	SIGNAL / NOISE	LOWEST POSSIBLE AT LIMIT OF RANGE	IN MANY CASES VERY HIGH (50 DB)
8.	DUTY CYCLE	RADIATION INTERMITTENT	RADIATION MOSTLY PERMANENT
9.	POWER LIMITATION	USUALLY PEAK POWER LIMITED	USUALLY AVERAGE POWER LIMITED
10.	BANDWIDTH OCCUPANCY	MOSTLY WIDE, SEVERAL MC/S	MANY TIMES SMALL, SEVERAL KC
11.	LINEARITY (OF XMTR)	NOT REQUIRED	EXTREMELY IMPORTANT
12.	CROSSTALK	UNIMPORTANT	EXTREMELY IMPORTANT

Fig. 1. Comparison of radar versus communications

range of operation. It is fairly high in many communication systems, where Gaussian circuit noise is frequently negligible against non-Gaussian external disturbances. Most radar transmitters operate intermittently, being designed for maximum peak power output, whereas most communication systems are designed for maximum average power output and operate continuously. Linearity is not a big issue in radar systems, but is of highest importance in most communication systems. Crosstalk between the channels of a multiplex system or between the elementary intervals of a digital system is a basic issue in communications, but is only a secondary problem in radar (separating closely spaced targets).

Such a large number of basic differences between radar and communications should be a warning against indiscriminate application to communication systems of those results of information theory which have been derived for the radar case. This applies particularly to the statement of radar detection theory, that the detectibility of a signal is primarily determined by the energy which a signal carries and will not be affected by the type of waveform in which this energy is being transmitted.

This statement is misleading when applied to the general model of a discrete communication system as shown in Fig. 2. A discrete information transmission system may be defined as one which transmits information by making discrete selections of predetermined signals at the transmitting side and which reconstitutes the information at the receiving side by making decisions about the type of signals, which most likely have been transmitted during a given time interval.

From this model it is seen that the main difference in a comparison with radar detection is the need to discriminate between different types of discrete signals. Thus there are two different problems which information theory has to solve for the communication system:

1. Optimum discrimination of each of the waveforms $f_1(t) \ldots f_n(t)$ against noise and disturbances, i.e., signal extraction.
2. Optimum decision between the extractor outputs, to determine which waveforms most likely have been transmitted during a given interval.

The first problem is the basic problem of radar detection theory. The second problem imposes additional requirements for specifying useful signals, and a very important one is the orthogonality requirement for admissible signals. They may all have the same energy, but good discrimination by the receiver requires that they have many different characteristics. Thus it cannot be claimed that only the signal energy determines the detectibility.

The term optimum must be explained. According to Middleton and van Meter,[1] ". . . optimization by theory can be carried out with respect to a criterion of excellence by which the performance of a reception system can be rated." (This criterion may be different for the design of radar systems and of communication systems.) "In practice, performance specifications are rarely explicit enough to fix the optimum system uniquely. More often, the designer must himself supply a definition of 'best' for the situation at hand, which is sufficiently precise to determine a unique system and which at the same time accurately reflects the given design constraints."

The foregoing definition of a discrete communication system and its detection problem still leaves the signal pattern undecided. Its only characteristic so far is that information will be exclusively expressed by selections out of predetermined waveforms.

The first specialization can be made by defining a digital system as a discrete communication system with (predeter-

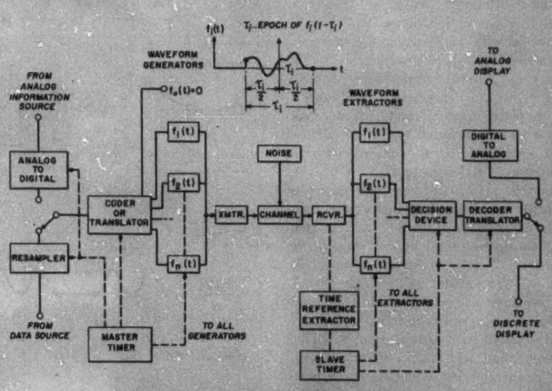


Fig. 2. Typical discrete communication system

mined) waveforms, which are limited in time and frequency. Any member $f_i(t)$ of a suitable ensemble of waveforms is a "symbol" of the signal alphabet. Each such symbol can then be nonzero only within the range

$$-\frac{T_i}{2} \leq t - \tau_i \leq +\frac{T_i}{2}$$

and must be essentially zero outside these limits. T_i is the digital interval for the symbol i and τ_i is the "epoch" of the symbol.

Following the definition of a digital system, one may further distinguish between single digital systems and multiple digital systems. The former never must send into the channel more than one signal at a time. The latter can tolerate the superposition of several signals, during the same interval or fraction of an interval. In Fig. 2 it is necessary (in single digital systems) to impose the constraint on the translator that no waveform must be triggered unless the digital interval of the previous waveform is at its end.

Any digital system may operate with the same interval length T and the same bandwidth occupancy W for all symbols, or it may have different WT values for the various symbols. The product WT may be called the signal base b of a symbol and one may speak of digital systems with uniform signal base, with staggered signal base, with weighted signal base, and with variable signal base. Systems with uniform signal base are preferred in channels with primarily Gaussian disturbances and whenever simplicity of the design is important. Staggered signal base is the same principle with a more refined matching of the signal base to the selection probability of the symbols or to the error probabilities of individual symbols in a channel with non-Gaussian disturbances. Variable signal base is applied preferably in channels with nonstationary disturbances, e.g., under jamming conditions.

Applications of information theory may now be demonstrated to the special case of a single digital system with uniform intervals. Referring to Fig. 2, a decision must be made as to how many symbols should be provided (number n), and then the actual functions $f_1(t)\ldots f_n(t)$, which all should be essentially zero outside the interval T and the frequency range W, must be prescribed.

Starting with n, the size of the signal alphabet, $f_0(t) = 0$ is included principally as an information-carrying symbol. This means that the state of nontransmitting, of no energy being submitted to the channel during specified intervals, is one of the predetermined symbols. Such a system is called a passive "$(n+1)$-ary" system, whereas a system which does not include this state may be called an active "n-ary" system. All its waveforms submit energy to the channel and thereby give an active indication to the receiver. An active system is automatically a "fail-safe" system, as any failure of the transmitter to send a symbol during any interval will be immediately registered at the receiver as a nonadmissible state. If the signal should fade during one or several intervals, the receiver will automatically locate such failures and will not mistake them for errors. This happens in the case of a passive alphabet, where a number of "space" digits would produce the same condition as an actual failure of the transmitter or a longer fade-out of the signals.

The detection problem in the passive system includes the hypothesis N, that noise alone is present during any interval. This hypothesis is nonadmissible in active systems. The passive binary system is actually the only one which can apply radar detection theory directly, with the only decision between the hypothesis N (noise alone being present) and the hypothesis SN (signal plus noise being present). The passive n-ary case has been theoretically investigated by the University of Michigan research team[2] under the assumption of a finite sampling plan and with the application to a coherent or noncoherent radar in mind, where the target is at one of a finite number of nonoverlapping positions. This case allows a decision device which can test first the hypothesis S_1N against N, then the hypothesis S_2N against N, etc., until S_nN against N, if $S_1, \ldots S_n$ are designated as the n-orthogonal signals.

The active n-ary case however excludes the hypothesis N completely and a multiple alternative decision of the hypothesis S_1N against S_2N and against all other S_iN until S_nN must be made. This decision problem has been covered by Middleton and van Meter[3] and their results may be applied for comparing the passive $(n+1)$-ary case with the active n-ary system. The above-mentioned advantage of the active system of offering automatic failure detection may be expressed by associating smaller cost $C_0^{(K)}$ with the decision that no signal was present, when actually signal K was transmitted. This cost ratio of failure against error depends on the chances for correcting failures, either by repetitions or by inspection of the output, against the chances for correcting errors, which are not yet located. It may be important to point towards the fact that in the passive binary case the energy of the only active signal can be twice the energy of either signal in the active binary case, provided both systems operate with the same average power and with equal probability for both transmission states. This power advantage of the passive system becomes smaller and smaller with increasing size of the signal alphabet. The passive $(n+1)$-ary system can use only $(n+1)/n$ times the energy per active element than the active n-ary system.

There is another subdivision of digital systems according to the size of the signal base b. If b (the WT product) is

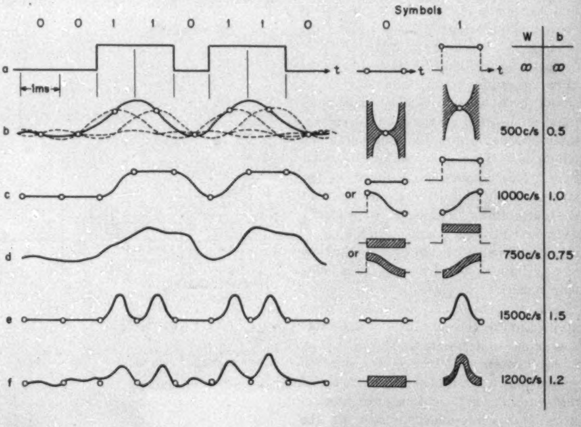


Fig. 3. Passive binary waveforms for data transmission

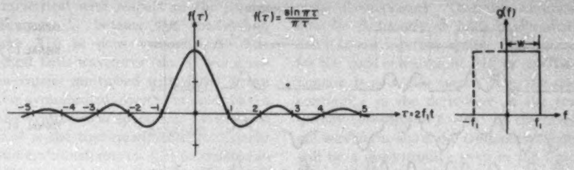


Fig. 4. Waveform (sin x)/x and its spectrum

$T=t_1=1/2f_1$, $W=f_1$, $WT=0.5$
$T_2=21/2f_1$, $W_2=f_1$, $W_2T_2=10.5$

relatively large, so that only a large number of independent samples will specify the waveform completely, statistical observations of signal and noise may be used for optimum detection systems, noiselike waveforms can be applied as signals and reliable detection of extremely weak signals in a noisy channel is possible.[3] Applying a large signal base in single digital systems leads either to a very small data rate per unit bandwidth or to extremely large alphabets.

The more frequent specification in communications calls for reliable data transmission over relatively good channels with the highest possible data rate. It is then preferably to operate with a very small signal base and possibly with large alphabets.

Before demonstrating some features of systems with larger alphabets, the author first discusses various forms of binary data transmission over a low-pass channel. It is interesting to see that ordinary teletype waveforms do not fall under the suggested definition of a digital system. Fig. 3 shows in waveform *a* a binary sequence 00110110, being represented by $f_0(t) = 0$ for 0 (space symbol) and by $f_1(t)$ = constant for *l* ("mark" symbol). This may be called a digital sequence in passive binary form and it may actually appear in this form at the output of a computor. Observe that the transitions between the two symbols are infinitely steep steps and that an infinitely wide band would be required theoretically to transmit this waveform. The signal base for the only active symbol specified would also go toward infinity.

For any practical transmission system it is desirable to arrive at the smallest possible signal base. Theoretical analysis assumes $(\sin x)/x$ sampling waveforms, whenever an attempt has to be made to specify a continuous function in terms of discrete parameters. Fig. 4 shows this waveform and its frequency spectrum. If one could define as the elementary interval the distance between any two adjacent zero crossings on either side of the center of the waveform, one would arrive at a signal base $b = 0.5$. But this definition is not in line with the one given previously for a digital system. The waveform would not be essentially zero outside such an interval. Insisting, for example, that the waveform should stay below 3% peak value everywhere outside the interval T would lead to a signal base $b = 10.5$.

Returning to Fig. 3, in waveform *b* is found the superposition of $(\sin x)/x$ signals in a passive binary system. Observe that the detection problem in this case is different, because one cannot specify the only admissible waveform over the whole interval. All that one may give to the receiver as a priori information about the waveform is its value at the center of the interval and a probability distribution for the waveform with increasing uncertainty towards the ends of the interval. At the limit, between two intervals, this uncertainty range goes even to plus and minus infinity, and this in the noiseless and distortionless case. This uncertainty results from the influence of the neighboring elements which, at least for the future, cannot be predicted. The optimum detector will in this case be a simple resampling device operating at the center of the waveform and a single threshold decider thereafter. If in addition to these theoretical arguments, the practical difficulties of generating and transmitting $(\sin x)/x$ waveforms are considered, it is evident that a signal base of 0.5 is unrealistic.

Waveform *c* shows a very economical solution, in which the whole interval is used for the transition between the two states. This transition follows purposely a positive or negative sine function from $-\pi/2$ to $+\pi/2$ with reference to the center of the waveform. It is seen that alternate transitions produce a permanent sine oscillation of a frequency equal to 1/2T. Putting this frequency at the center of the low-pass band will permit a practically distortion-free transmission of binary data at a rate equal to one bit per cycle bandwidth. The signal base *b* is unity. Observe, however, that again the waveforms $f_0(t)$ and $f_1(t)$ cannot be specified uniquely. It is seen that space can either be transmitted by a constant zero level or by a negative sine transition from peak level to zero level; similarly, for the mark symbol. Only at the end of the interval does the transmitter reach a unique state, which does not depend on the information of the previous element. This system cannot be considered as a binary digital system. It could better be designated as an analog transmission of binary information. It falls under the class of discrete communication systems, but not under the subclass of digital systems.

There is however the possibility of considering this system as a passive quaternary digital transmission system with a rate of one bit per cycle bandwidth over a low-pass channel and with four completely specified waveforms of which two represent space and two mark. The waveforms within each subset indicate also the symbol of the preceding interval, which fact may be used for error-checking procedures. This leads to the advantages of digital systems with larger signal alphabets.

Waveform *d* demonstrates the rapidly increasing uncertainty of anyone of these four waveforms, when reducing the signal base below unity. This uncertainty is again caused by intersymbol crosstalk and may be called self-noise of the system. It will be useless to reduce the signal base so much that the self-noise will approach the external noise level. The increase of signaling rate would be compensated by the increase in equivocation and the effective information transmission rate would not be increased.

Waveform *e* finally shows the only truly binary digital data transmission system on this slide. It is still a passive binary system, but a waveform $F_1(t)$ for the transmission state mark may be specified uniquely. This waveform will be submitted to a low-pass channel, whenever mark has to be transmitted, regardless of the information in any other intervals. Observe that another necessary condition for a digital system is that the selected waveforms have to start and to end all in the same transmission state, here at zero level. This will avoid any transients between the intervals, whatever combination of symbols might be selected. It may be seen that any possible transition, e.g., 00, 01, 10 and 11, can now be performed without appreciable intersymbol crosstalk. This advantage is paid for with a signal base of about 1.5, i.e., only a bit rate of about two thirds of the low-pass bandwidth can be

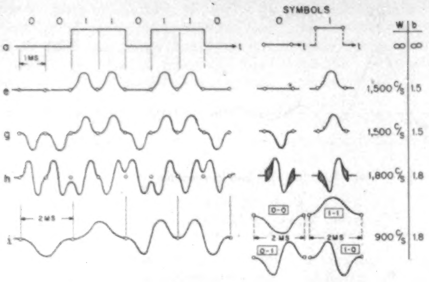
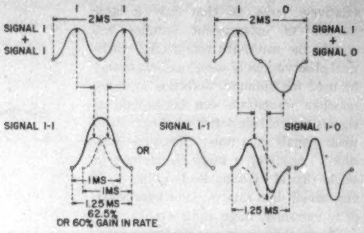


Fig. 5. Active binary and quaternary waveforms



THE SIGNALS 0-0 AND 0-1 ARE FORMED BY THE NEGATIVES OF ALL THE ABOVE WAVEFORMS.

Fig. 6. Forming quaternary signals by overlapping binary signals

achieved. However the advantage of optimum detection systems is gained. Again the signal base may be reduced to about 1.2 and a certain uncertainty range due to ringing effects following the active signals may be tolerated, as demonstrated in waveform *f*.

When applying optimum detection theory to this case of very small signal base (*b* close to 1), a particular sampling plan cannot be applied, operating with $2WT$ independent samples. Middleton and van Meter[1] calculate, in Appendix I of their decision theory, the error which results from the missing contributions of samples outside the observation interval and conclude that the signal base should be at least ten to permit the neglecting of these fringe effects of a finite sampling plan. For a very small signal base, which is the rule in most communication problems, they propose an alternative procedure, leading to continuous sampling as a limiting case of discrete sampling with the number of samples increasing to infinity and the observation interval T fixed. This leads apparently to the application of matched filters or continuous cross-correlators.

So far it is seen that the desire of applying an optimum detection system leads to a definition of digital systems, which requires uniquely specified waveforms. These again require a signal base of at least 1.2. The analytical form of signals with a minimum signal base must still be defined. Since the uncertainty principle prevents the simultaneous specification of time and frequency with absolute accuracy, the limits for the bandwidth W and the time T must be defined on a tolerance basis. This can be done by insisting that the percentage of the whole energy of the waveform outside the limits W and T should be less than a certain small value, or by specifying that the maximum amplitude and maximum spectral density of the waveform outside the limits W and T should be less than a certain small fraction of the peak value of these magnitudes inside these limits.

Fig. 5 shows in line *a* the same binary sequence as in the previous slide. Below this, there is a waveform representing the passive binary system using $\cos^2$ signals. This is compared in line *g* with an active binary system, using positive and negative $\cos^2$ signals. Finally in line *h* an active binary system with waveforms of more complex character is shown. Space is represented by a waveform consisting basically of one period of a negative sine oscillation and mark is represented by a similar waveform of a positive sine oscillation.[4,5] Both require a signal base of about 1.8. Adjacent waveforms will still slightly overlap, causing a small uncertainty range at the ends of the interval. The increased signal base of this system reduces the signaling rate per unit bandwidth, but each signal now carries slightly more energy for the same peak-to-peak voltage. This increases the detectibility. These signals have also the additional advantage that they do not possess a d-c component. Thus there will be no variation of the base line depending upon the information when using these waveforms.

Combining lines *g* and *h* into a single system may be a first step toward a higher order alphabet. It is seen in line *i* how two binary elements together, i.e., binary diagrams, can form a quaternary information symbol. There are four such symbols, 00, 01, 10, and 11. Each is represented by a uniquely specified waveform. All these waveforms are now twice as long as before, require therefore only half of the bandwidth, but each waveform carries two bits of information. The quaternary system of line *i* has only half of the digital rate of the binary system of line *g*, but has the same bit rate. Its advantage is the smaller bandwidth required for the same bit rate and signal power. Smaller bandwidth means also smaller noise power, thus an improvement in input signal-to-noise ratio in the receiver.

Fig. 6 demonstrates the reason for this improvement when going to a higher order alphabet. It can be shown that the four quaternary waveforms result from superposition of the respective binary waveforms. In the quaternary system the two waveforms can be overlapped at the center of the quaternary interval and thus the quaternary interval can be reduced about five eighths of the two binary intervals. This is the real gain of the higher order alphabet. Extending this method, all 32 combinations of the five binary elements could be specified in the teletype code and form a system with an alphabet of 32 waveforms, with a correspondingly large saving in bandwidth and with considerably increased detectibility, as it would be a truly digital system. However, it would not be an optimum alphabet of waveforms.

Several publications of information theory[6] stressed the fact that an optimum set of waveforms, matched to a set of optimum signal extractors, should be orthogonal. The waveforms in such a set have to satisfy the condition

$$\int_{-T/2}^{+T/2} f_{i(t)} \times f_j dt_{(t)} = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}$$

The second condition makes the set normal.

A set of waveforms, which contains all possible waveforms of a certain class and which, can be orthogonal to each other in a given interval T, is called a closed orthogonal set. In carrier bands an attempt shall be made to find closed sets of normal and orthogonal waveforms by extending the definition to complex functions. Wiener[7] shows one possible method to derive such sets.

This leads to the last point to be mentioned. It seems that in digital transmission systems it is better to specify directly the set of waveforms within the carrier band. One may speak of pass-band waveforms and it seems that it is better to go to a larger alphabet directly within the carrier band, in place of defining a smaller alphabet within the low-pass band and transferring it by a modulation process to the carrier band.

There are good reasons for such a philosophy: it can be shown that the signal base increases when going from the low-pass band to the corresponding band-pass channel. If this is done in double side-band (DSB) modulation, the bandwidth doubles and the interval remains the same. If there is a transfer into single side-band (SSB) modulation, the bandwidth remains the same, but the interval increases. It is easy to show (see the Appendix) that the $(\sin x)/x$ waveform has in SSB an envelope with twice the distance between adjacent zero crossings, compared with the modulating low-pass waveform. Other SSB waveforms show a corresponding increase of the signal base.[8] Notwithstanding this increase of the signal base, it is found that the information-handling capacity remains constant, because in the band-pass channel there are two independent components for each waveform. They are usually called the quadrature components; using them as independent symbols doubles the size of the signal alphabet.

Fig. 7 shows as an example how one of the skew-symmetrical wave forms, which were used before in the quaternary system, is first converted to two different band-pass waveforms in DSB with suppressed carrier. Observe that the right side of the slide shows the Fourier spectrum of all these waveforms as a real function of frequency $g(f)$ ranging from $-\infty$ to $+\infty$. This completely real form of Fourier transform has been closely investigated by Hartley[9] and is called the Hartley spectrum to discriminate it from the usual complex Fourier spectrum or from the Fourier spectrum with the two real quadrature conponents (sine and cosine spectrum). It has the characteristic that symmetrical waveforms have a symmetrical spectrum and skew-symmetrical waveforms have a skew-symmetrical spectrum.

The DSB waveform, no. *2*, has a sine carrier multiplied with a skew-symmetrical waveform $A(t)$. This gives a symmetrical product function, as indicated by the actual waveform at the center. The Hartley spectrum is therefore also symmetrical with respect to $f = 0$. The two parts of the spectrum are both skew symmetrical with respect to the carrier frequency f_c, because the modulating wave $A(t)$ is skew symmetrical. The second DSB waveform (no. *3*) has a cosine carrier multiplied with $A(t)$. It is a skew-symmetrical waveform and has a skew-symmetrical spectrum. Waveform no. *4* is the time-quadrature component (Hilbert transform) of $A(t)$ modulated in DSB to a cosine carrier. The superposition of this waveform with waveform no. *2* results in the phasing method of SSB modulation. It is seen in the last line that the envelope of the resulting SSB waveform is much wider than the interval of the original function. The envelope is, as demonstrated in the Appendix of this paper, the root of the sum of the squares of the modulating function and its Hilbert transform. The SSB wave will in general have a phase-modulation term which is the arc-tangent of the quotient of the Hilbert transform over the modulating function.

It is very significant to follow this process in the Hartley spectrum. This spectrum contains the full information of each waveform, including the epoch (origin of the time scale). On each half of the spectrum of waveform no. *4* is seen the low-pass spectrum of the Hilbert transform of waveform no. *1*. This can be easily designed graphically, by merely taking the negative-half spectrum of line 1 with negative sign. If the original Hartley spectrum is neither symmetrical nor skew symmetrical, it must first be split into a symmetrical and skew-symmetrical component. Then this operation may be performed on both components and it is put together again. The reason for the wide extension of $A(t)$ in the time domain is now also seen. It is the discontinuity in the derivative of the frequency spectrum at $f = 0$. If the original waveform has a d-c component there will be a discontinuity even in the function $g(f)$ itself. Since frequency spectrum and time function are interchangeable, this spectrum may be considered as a time function, and every engineer knows that a waveform with such a sharp corner will produce a wide frequency band. Thus one understands that for SSB transmission one should use waveforms without a d-c component and possibly with horizontal tangent of the spectral curve at d-c.

However, the better approach appears to be to ask directly for the carrier-band waveform which may fit best into the given SSB and not to go through the complicated process over two modulators, one wide-band phase shifter and one combiner. If it is intended to find a waveform which would have its energy uniformly distributed over the given frequency band (SSB), one arrives at waveforms tentatively called full-band waveforms. Fig. 8 shows the two quadrature components of such full-band waveforms and their spectra. The Appendix contains the short mathematical proof of the emergence of these waveforms, by one of three modulation methods:

1. Modulation of a $(\sin x)/x$ waveform in

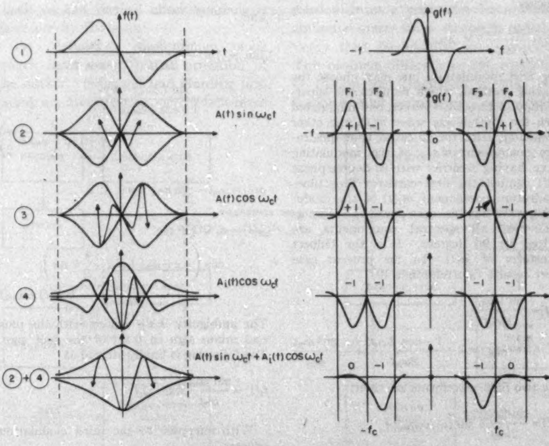


Fig. 7. SSB waveform generation and Hartley spectrum

DSB suppressed carrier modulation to a carrier at the center of the given band.

2. Modulation of a (sin $2x$)/$2x$ waveform in SSB to a carrier at the lower end of the band, selecting the upper side band in this process.

3. Modulation of a (sin $2x$)/$2x$ waveform in SSB to a carrier at the upper end of the band, selecting the lower side band.

In all three cases the result is identically the same full-band waveform with a (sin x)/x envelope and a constant instantaneous frequency at the center of the band showing no phase modulation.

At least in this case of the full-band waveforms, full identity is found between SSB and DSB waveforms. It appears therefore logical not to pay too much attention to the modulation system, but more to the carrier-band waveforms directly in digital transmission systems of high bit rate per cycle bandwidth.

Appendix

It shall be proved that the attempt to transmit a (sin $\omega_m t$)/$\omega_m t$ signal in amplitude modulation of a carrier, i.e., cos $\omega_c t$, leads identically to the same band-pass waveform (modulated signal), whether the modulation process is:

1. DSB suppressed carrier modulation, modulating ω_c with (sin $\omega_m t$)/$\omega_m t$, or
2. SSB suppressed carrier modulation of $\omega_c - \omega_m$, selecting the upper side band, and modulating with (sin $2\omega_m t$)/$2\omega_m t$, or
3. SSB suppressed carrier modulation of $\omega_c + \omega_m$, selecting the lower side band and modulating with (sin $2\omega_m t$)/$2\omega_m t$.

With reference to the first modulation process

$$f(t) = \frac{(\sin \omega_m t)}{\omega_m t} \times \sin \omega_c t$$

For SSB modulation, one may choose the phasing method, which requires the superposition of two DSB waves, one modulated with the modulating wave $m(t)$, the other modulated with the so-called time-quadrature component $m_t(t)$ of the modulating wave, having a carrier with 90-degree phase shift against the first carrier. This time-quadrature component $m_t(t)$ of the modulating wave has the same power spectrum as $m(t)$, but all spectral components are shifted for 90 degrees. It is the Hilbert transform of $m(t)$. In the present case there results (see reference 10)

$$m(t) = \frac{\sin 2\omega_m t}{2\omega_m t};$$

$$m_t(t) = -\frac{1 - \cos 2\omega_m t}{2\omega_m t} = -\frac{\sin^2 \omega_m t}{\omega_m t}$$

The two SSB waveforms are then:

$$f_1(t) = \frac{\sin 2\omega_m t}{2\omega_m t} \sin (\omega_c - \omega_m)t + \frac{1 - \cos 2\omega_m t}{2\omega_m t} \times \cos (\omega_c - \omega_m)t$$

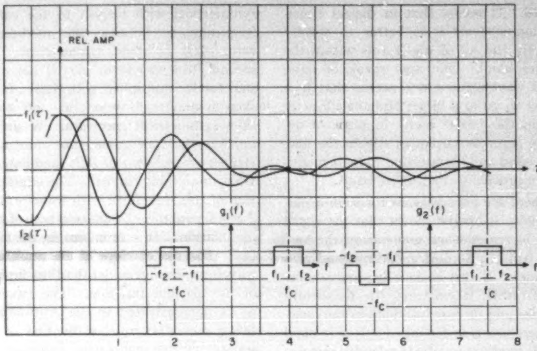


Fig. 8. Full-band waveforms

$$f_2(t) = \frac{\sin 2\omega_m t}{2\omega_m t} \sin (\omega_c + \omega_m)t - \frac{1 - \cos 2\omega_m t}{2\omega_m t} \times \cos (\omega_c + \omega_m)t$$

With reference to the second modulation process

$$f_1(t) = E_1(t) \cos [(\omega_c - \omega_m)t + \phi(t)]$$

with

$$E_1(t) = \pm\sqrt{\left(\frac{\sin 2\omega_m t}{2\omega_m t}\right)^2 + \left(\frac{1 - \cos 2\omega_m t}{2\omega_m t}\right)^2} = \pm\sqrt{\frac{2 - 2\cos 2\omega_m t}{4\omega_m{}^2 t^2}}$$

$$E_1(t) = \pm\sqrt{\frac{4 \sin^2 \omega_m t}{\omega_m{}^2 t^2}} = \frac{\sin \omega_m t}{\omega_m t}$$

and

$$\tan \phi(t) = -\frac{\sin 2\omega_m t}{2\omega_m t} \times \frac{2\omega_m t}{1 - \cos 2\omega_m t} = +\frac{\sin (2\omega_m t - \pi)}{1 + \cos (2\omega_m t - \pi)} = +\tan\left(\omega_m t - \frac{\pi}{2}\right)$$

$$\phi(t) = \omega_m t - \frac{\pi}{2} \pm n\pi, \quad n = 0, 1, 2 \ldots$$

$$f_1(t) = \pm\sqrt{E_1{}^2(t)} \times \cos\left(\omega_c t - \omega_m t + \omega_m t - \frac{\pi}{2} \pm n\pi\right) = E(t) \sin \omega_c t$$

The ambiguity $\pm n\pi$ cancels with the plus and minus sign in front of the root sign. The following is finally arrived at

$$f_1(t) = \frac{\sin \omega_m t}{\omega_m t} \times \sin \omega_c t$$

With reference to the third modulation process

$$f_2(t) = E_2(t) \cos [(\omega_c + \omega_m)t + \phi(t)]$$

$$E_2(t) = E_1(t) = \frac{\sin \omega_m t}{\omega_m t}$$

$$\tan \phi(t) = \frac{\sin 2\omega_m t}{2\omega_m t} \times \frac{2\omega_m t}{1 - \cos 2\omega_m t} = -\frac{\sin (2\omega_m t + \pi)}{1 + \cos (2\omega_m t + \pi)} = -\tan\left(\omega_m t + \frac{\pi}{2}\right)$$

$$\phi(t) = -\omega_m t - \frac{\pi}{2} \pm n\pi$$

$$f_2(t) = \pm\sqrt{E_2{}^2(t)} \times \cos\left(\omega_c t + \omega_m t - \omega_m t - \frac{\pi}{2} \pm n\pi\right) = E_1(t) \sin \omega_c t$$

Again it is seen that the ambiguity of the plus and minus signs cancels and the result is

$$f_2(t) = \frac{\sin \omega_m t}{\omega_m t} \times \sin \omega_c t = f_1(t) = f(t)$$

It may be seen also that the envelope in SSB is (sin $\omega_m t$)/$\omega_m t$, when the modulating wave is (sin $2\omega_m t$)/$2\omega_m t$. Thus the distance between any two adjacent zero crossings on the time axis doubles, when going from the modulating wave to the envelope of a SSB wave.

This proof demonstrates clearly that there is only one type of full-band wave, no matter which modulation process is chosen. However, there are the two phase-quadrature components of this wave. The same equations can be derived for

$$\frac{\sin \omega_m t}{\omega_m t} \times \cos \omega_c t$$

which is orthogonal to

$$\frac{\sin \omega_m t}{\omega_m t} \times \sin \omega_c t$$

The total band for any of these waves ranges from $\omega_c - \omega_m$ to $\omega_c + \omega_m$. The bandwidth is $2\omega_m$. Since there are two phase-quadrature components, four $\omega_m t$ orthogonal waveforms can be found again within the band $2\omega_m$ and the time T by placing the peaks of each pair of quadrature waveforms into the zeros of each other pair of such full-band waveforms.

References

1. Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory, **D. Middleton, D. van Meter.** *Journal*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., vol. 3, no. 4, Dec. 1955, pp. 192–253; vol. 4, no. 2, June 1956, pp. 86–119.

2. The Theory of Signal Detectibility, **W. W. Peterson, T. G. Birdsall, W. C. Fox.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., vol. 4, Sept. 1954, pp. 171–212.

3. On Optimum Multiple-Alternative Detection of Signals in Noise, **D. Middleton, D. van Meter.** *Ibid.*, vol. 1, Sept. 1955, pp. 1–9.

4. Anti-jamming and Security Performance of Passive Ternary Data Transmission of Binary Signals, **R. Filipowsky, E. Scherer.** *Technical Report no. 4*, Communications Engineering, Electronics Division, Westinghouse Electric Corporation, Baltimore, Md., 1956.

5. Frequency Spectra and Transients of Some Skew-Symmetrical Elementary Signals (Swinging Ternary Signals), **E. Scherer.** *Technical Report No. 2*, Communications Engineering, Electronics Division, Westinghouse Electric Corporation.

6. An Anti-Multipath Communication System, **R. Price, P. E. Green.** *JA-787*, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., Aug. 1957, p. 12; also Bibliography, p. 207.

7. The Extrapolation, Interpolation and Smoothing of Stationary Time Series (book), **Norbert Wiener.** John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 31–36.

8. Pulses and Transients in Communication Circuits (book), **C. Cherry.** Chapman & Hall, Ltd., London, England, 1949, chap. 7. "Distortion of Rectangular Pulse" in SSB envelope.

9. A More Symmetrical Fourier Analysis, **R. V. L. Hartley.** *Proceedings*, Institute of Radio Engineers, vol. 30, Mar. 1942, pp. 144–50.

10. Information Theory (book), **S. Goldman.** Constable & Company, London, England, 1953, App. V.

Optimum Block Length for Data Transmission with Error Checking

F. B. WOOD
ASSOCIATE MEMBER AIEE

Synopsis: The efficiency of a block-checking system of data transmission, in which an answer-back block-checking signal is provided to acknowledge each block of characters, or to ask for a repeat if an error is detected, is investigated. Curves of data transmission efficiency, as a function of the number of characters in a message block, are shown to have sharper peaks as the distance and error probabilities are increased. For comparison, efficiency curves are plotted for a 7-bit single-error-detecting code and for a 10-bit single-error-correcting code. A family of curves of optimum block length is given for different probabilities of error and different reply times.

A BLOCK-CHECKING SYSTEM is one in which there is a periodic feedback signal which acknowledges correct receipt of the last block of *n* characters or asks for a repeat of the last block, as illustrated in Fig. 1. A block-check symbol is added after the *n*th character of the message block. It may be a simple end-of-block signal, or it may contain a parity count of the bits in particular rows of the message. To distinguish between different types of parity checks, a message can be visualized as a matrix where each column contains the bits of a single character. Then the addition of a parity bit to each character constitutes a vertical check. The block-check symbol then becomes a longitudinal check which can consist of the parity bits for each row. A special case of a block-check symbol providing a longitudinal check on the vertical check symbols is described by Barbeau.[1]

Two definitions of optimum block length have come to the attention of the author. Schatzoff and Harding proposed an investigation of the optimum block length for minimizing the undetected errors.[2] A. Cohen, in an unpublished report on transmission of punched card data over teleprinter lines, defined optimum block length as the number of characters per block which would maximize the transmission efficiency. Cohen's definition of optimum block length is used in this study.

The transmission efficiency is the time required to transmit the information bits divided by the total time required to transmit the information bits plus the time required for redundancy bits, waiting time for the answer-back signals, and the average additional time for repeated transmissions.

Bishop and Buchanan have proved a theorem that redundancy for error correction is more economical than using feedback to request a repeat when an error occurs.[3] However, automatic transmission from a computer to a remote station requires some feedback signal to verify that the channel is satisfactory. This problem differs from the conditions of Bishop and Buchanan in that both feedback and redundancy are required

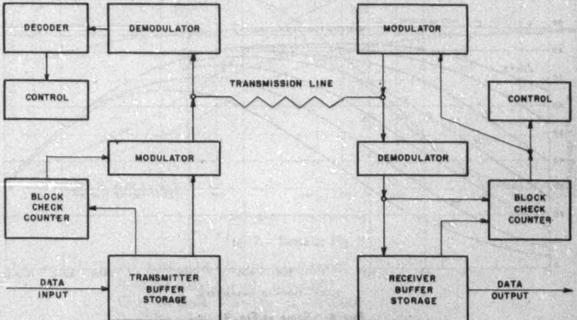


Fig. 1. Block-check answer-back system

Paper 58-1181, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 13, .1958; made available for printing July 31, 1958.

F. B. Wood is with International Business Machines Corporation, San Jose, Calif.

The author wishes to acknowledge the assistance received from Dr. W. A. Christopherson in formulating the problem. T. C. Kelley assisted in the programming of the *650* computations. Dr. A. B. Fontaine made valuable suggestions for simplifying the mathematical derivations. Dr. G. C. Preston and Dr. R. C. Wrede gave valuable assistance through numerous discussions of probability theory.

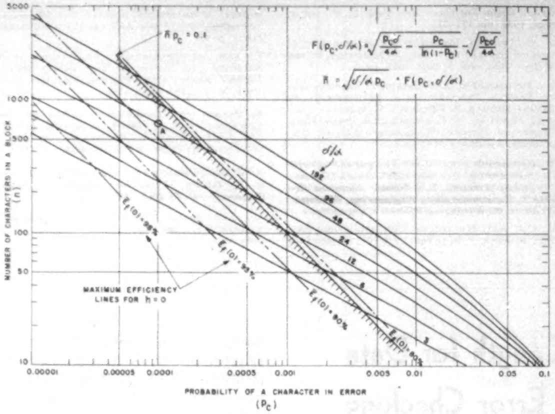


Fig. 2. Optimum block length versus probability that a character is in error, loci of maximum efficiency

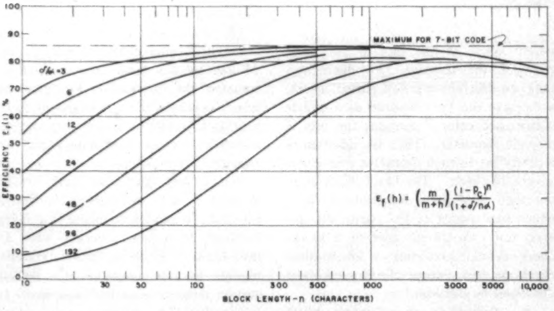


Fig. 3. Data transmission efficiency for specified error reply time δ/α and probability of character error p_0 as function of block length n

$h = 1$, $p_0 = 0.00001$

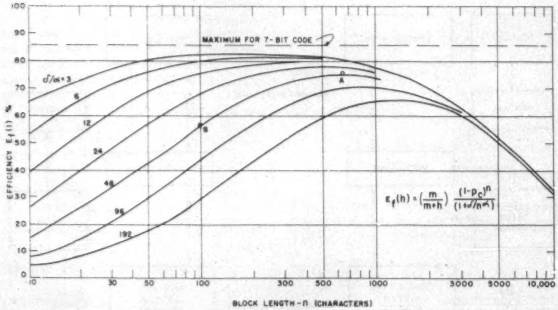


Fig. 4. Same as Fig. 3

$h = 1$, $p_0 = 0.0001$

so that the question is to determine the most economical ratio between redundancy and feedback for a particular set of conditions.

The effect of line failure when the automatic switching system fails to find a spare channel is not included in this analysis. For data on the probability of such events, the reader is referred to reports of Bell Telephone Laboratories.[4]

An answer-back signal is a signal sent by the receiving station to indicate acknowledgment or to request a repeat of the last block. When no errors are detected, the net information transmission rate is reduced by the time involved in coding the block-check symbol, propagation time over long line, decoding logic time, and setting of telephone line echo suppressors. When errors are detected, the net transmission rate is further reduced by the time required to repeat blocks containing errors.

Efficiency is plotted for some sample conditions of error probability and total block-check and answer-back time. A general equation for optimum block length is derived by setting the derivative of the efficiency with respect to the number of characters in a block equal to zero to obtain the condition of maximum efficiency. A set of curves is plotted for different values of reply time and character error probability.

The general equation for optimum block length reduces to Cohen's equation when the product of the block length and the probability of error are small. Sample values of optimum block length and efficiency are plotted for a range of error probabilities and transmission speeds.

The curves in this report can be used to determine the optimum block length and efficiency in practical cases by the following steps:

1. Calculate time interval per character.
2. Calculate the reply delay time interval required for the block-check and answer-back signals, and determine the number of characters which could be sent in the reply delay interval.
3. Select the curve in Fig. 2 corresponding to the number of characters determined in item 2, then determine from statistical data the probability that a character is in error. The intersection of the character error probability with the optimum block length curve corresponding to the number of characters per delay time interval gives a point which graphically yields two items:
   a. The ordinate gives the optimum block length in the number of characters per block.
   b. A projection from the point parallel to the diagonal lines shows the maximum

efficiency that can be obtained. These diagonal lines are computed for zero redundancy, so they must be multiplied by the ratio of information to total bits per character.

4. To determine the efficiency at a block length other than the optimum value use equation 13 or read the value off the efficiency curves of Figs. 3 through 6.

A sample calculation is included in Appendix III for the foregoing steps. The use of these curves permits a comparison of line rental costs with buffer storage costs. It also is of use for comparison with error correction codes. Curves for a 10-bit single-error-correcting code are included in Fig. 7 for comparison with the more frequently used 7-bit single error-detecting codes.

The derivation of the efficiency exhibits an interesting phenomenon of different probability functions for the two levels of treatment of the problem. The probability of different numbers of characters being in error within a block follows the binomial distribution. When the probability of a block of n characters being correctly transmitted is considered, the probability follows the Pascal distribution.

Efficiency of Transmission for Single Error-Detecting Code

DEFINITIONS

The efficiency of transmission E_f is defined as the time required to transmit the information bits (excluding redundancy bits) of a block of n characters without any error-detecting or error-correcting system, divided by the average time required to transmit a block of n characters with all detected errors corrected. Where T is the average time to transmit a block of n characters with all single and multicharacter errors (detectable by the code used) corrected, T_1 is the time to transmit a block without correcting the errors, and each character contains m information bits and h redundant bits, the foregoing definition results in the following:

$$E_f(h)=\left(\frac{T_1}{T}\right)\left(\frac{m}{m+h}\right)\times 100 \qquad (1)$$

Where α is the time to transmit one character, and δ is the time between the end of the last character in a record and the beginning of the first character in the next record or the next repeat transmission

$$T_1=n\alpha \qquad (2)$$

and a single block with no detected errors, but with a feedback acknowledgment signal, takes a time interval

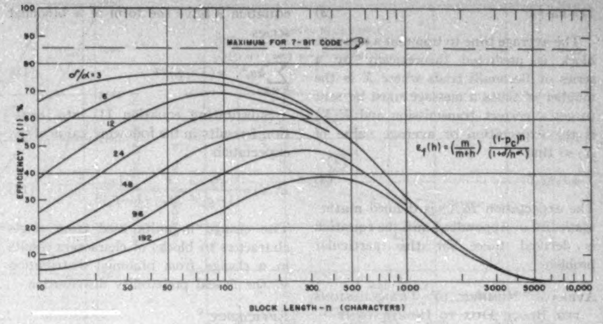


Fig. 5. Same as Fig. 3

$h=1$, $p_e=0.001$

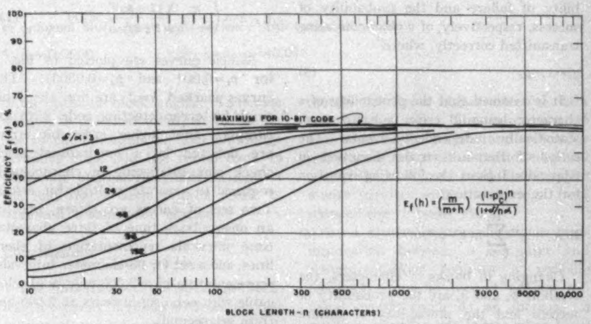


Fig. 6. Same as Fig. 3

$h=4$, $p_e=0.001$; $p_e''\sim p_e^2=0.000001$

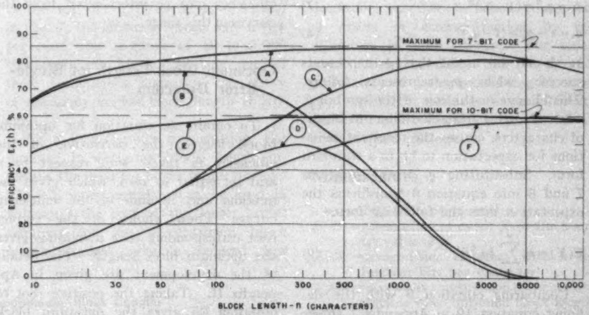


Fig. 7. Same as Fig. 3

A—7-bit code, $\delta/\alpha=3$; $p_e=0.00001$
B—7-bit code, $\delta/\alpha=3$; $p_e=0.001$
C—7-bit code, $\delta/\alpha=96$; $p_e=0.00001$
D—7-bit code, $\delta/\alpha=96$; $p_e=0.001$
E—10-bit code, $\delta/\alpha=3$; $p_e=0.001$; $p_e''\sim 0.000001$
F—10-bit code, $\delta/\alpha=96$; $p_e=0.001$; $p_e''\sim 0.000001$

$$T_0 = n\alpha + \delta \tag{3}$$

The average time to transmit a message block is predicted theoretically by a series of Bernoulli trials where X is the number of times a message must be sent to get a correct transmission and $E(X)$ is the expectation or average value of X, so that

$$T = T_0 E(X) \tag{4}$$

The expectation $E(X)$ is defined mathematically in Appendix I and the equation is derived there for this particular problem.

AVERAGE NUMBER OF TRANSMISSIONS PER BLOCK DUE TO INDEPENDENT STATISTICAL CHARACTER ERRORS

The symbols p_c and q_c are the probability of failure and the probability of success, respectively, of a character being transmitted correctly, where

$$q_c = 1 - p_c \tag{5}$$

It is assumed that the probability of a character being in error is based upon statistically independent events. The series of Bernoulli trials discussed in Appendix I gives the following equation for the expectation

$$E(X) = q_c^n \sum_{k=1}^{\infty} k(1-q_c^n)^{k-1} \tag{6}$$

Changing to blocks of characters, the symbols p_s and q_s are the probability of success and the probability of failure respectively, in transmitting a correct message of n characters. The relationship between the two probabilities is

$$p_s = q_c^n = (1-p_c)^n \tag{7}$$

$$q_s = 1 - p_s = (1-q_c)^n \tag{8}$$

It should be noted that p_s represents success, while p_c represents failure. This change in the role of the symbol p, when changing from characters to blocks of characters, causes the resultant equations for expectation to be in a standard form. Substituting q_c from equations 7 and 8 into equation 6 transforms the expectation into the following form:

$$E(X) = p_s \sum_{k=1}^{\infty} k q_s^{k-1} \tag{9}$$

Comparing equation 9 with the defining equation 19 in Appendix I shows that the probability function is of the form

$$P(X=k) = q_s^{k-1} p_s \tag{10}$$

which is recognized as the Pascal distribution described by Feller.[6] The terms under the summation sign of equation 9 have the form of a binomial series

$$\sum_{k=1}^{\infty} k q_s^{k-1} = (1-g_s)^{-2} \tag{11}$$

Substituting equation 11 into equation 9 results in the following value of the expectation

$$E(X) = \frac{p_s}{(1-q_s)^2} = \frac{1}{p_s} = \frac{1}{(1-p_c)^n} \tag{12}$$

The change in units used from single characters to blocks of characters results in a change from binomial distribution to the Pascal probability distribution.

EFFICIENCY

Using equations 1, 2, 3, 4, and 12 gives a transmission efficiency of

$$E_f(h) = \left(\frac{m}{m+h}\right)\frac{(1-p_c)^n}{(1+\delta/n\alpha)} \times 100 \tag{13}$$

Sample curves are plotted in Fig. 7 for $p_c = 0.001$ and $p_c = 0.00001$. The curves marked $h=1$ are for the 7-bit single-bit error-detecting code with assumption that higher order bit errors are detected by a longitudinal block check plus interlacing of characters if required to separate multiple-bit errors. Two sets of curves are shown, one for an answer-back time of three character time intervals representative of short lines, and a set for 96-character intervals, representative of a 3,000-mile nonloaded cable with echo suppressors at 2,000 bps (bits per second).

Additional curves for a range of δ/α for $p_c = 0.001$, 0.0001, and 0.00001 are plotted in Figs. 3 through 5 so that the efficiency can be interpolated from the curves in this range.

Optimum Block Length for Single-Error Detection

To obtain an equation for optimum block length, the derivative of the efficiency is taken with respect to n and set equal to zero, which gives the maxima and minima of the efficiency curves. Then, choice of the correct root corresponding to a maximum gives the optimum block length. The details of the development are given in Appendix II. Taking the positive root of equation 25 gives the optimum block length

$$\bar{n} = \sqrt{\delta/\alpha p_c}\, F(p_c, \delta/\alpha) \tag{14}$$

$$F(p_c, \delta/\alpha) = \sqrt{[1/S(p_c)] + (\delta/\alpha)p_c/4} - \sqrt{(\delta/\alpha)p_c/4} \tag{15}$$

When $p_c < 1/(100\ \delta/\alpha)$, which may also be stated as $\bar{n}p_c < 0.1$, $F(p_c, \delta/\alpha)$ approaches unity, so that the following equation has less than 5% error:

$$\bar{n}' = \sqrt{\delta/\alpha p_c} \tag{16}$$

Curves of optimum block length are plotted in Fig. 2 for different reply times (δ/α). The complete formula of equation 14 must be used for $\bar{n}p_c > 0.1$. Equation 16 can be used for $\bar{n}p_c < 0.1$.

Lines of equal efficiency have been added to Fig. 2. Substituting $\bar{n}$ in equation 13 and expanding the $(1-p_c)^n$ term in equation 13 by the binomial expansion, and dropping second-order and higher terms of p_c gives the following equation for maximum efficiency with less than 0.5% error when $\bar{n}p_c < 0.1$:

$$\bar{E}_f(h=0) = \frac{\bar{n} - \delta/\alpha}{\bar{n} + \delta/\alpha} \tag{17}$$

Efficiency for Single-Error Correction Code

For comparison of single-error detection codes ($m=6$, $h=1$) with an answer-back signal with a single-error correction ($m=6$, $h=4$) with answer-back signal, the efficiency of the error-correcting code is needed. In this sense "single error" means a single-bit error within a character. Therefore no repeats will be required for the single errors, but repeats will be asked for in the case of multiple errors detected by the addition of longitudinal block checking. Substituting the p_c'' derived in Appendix III for the 10-bit single-error correcting code of the Hamming[6] type into equation 13 gives

$$E_f(h=4) = \frac{m}{m+h} \times \frac{(1-45p_b^2(1-p_b)^8 - \ldots)^n}{(1+\delta/n\alpha)} \simeq \frac{m}{m+h}\,\frac{(1-p_c^2)^n}{(1+\delta/n\alpha)}, \text{ for } p_c \leq 0.001 \tag{18}$$

Examples of Optimum Block Length and Efficiency

Examples have been taken graphically from Fig. 7 to obtain the optimum block length and efficiency for a range of bit rates between 75 bps and 6,000 bps. These examples were based on nonloaded side circuits on a 3,000-mile 19-gage cable. These results are plotted (see Fig. 8) to show the change in optimum block length and maximum efficiency.

Curves have been plotted for lines with and without echo suppressors. Echo suppressors are used on most voice

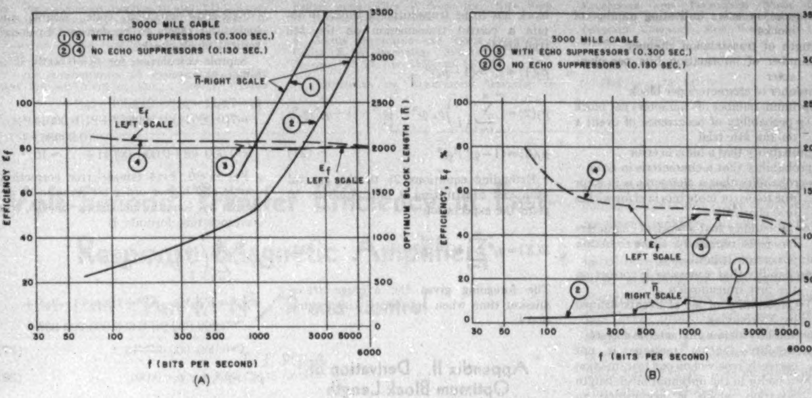


Fig. 8. Variation of optimum block length and efficiency with transmission rates:
A—$h=1$, $p_e=0.00001$ B—$h=1$, $p_e=0.001$

telephone circuits of over 500 miles. It is common to remove the echo suppressors from leased lines used solely for data transmission. A brief description of the functioning of echo suppressors may be found in a text issued by American Telephone and Telegraph Company.[7] Examination of Fig. 8 shows that the presence of echo suppressors only reduces the maximum attainable efficiency a few per cent, but their presence raises the optimum block length by as much as 50% for low-error probabilities such as $p_e=0.00001$.

This leads to the consideration of the case of a fixed buffer of less than the optimum block length. Some sample values of $\bar{n}$ and E_f are listed in Table I. If $p_e=0.001$ and $\delta/\alpha=24$, the optimum block length of $\bar{n}=130$ would give an efficiency of 69%, while the use of a buffer holding 100 characters would yield an efficiency of 62.5%. If the error probability decreases, the maximum efficiency goes up to 83% at the optimum block length of 1,300 characters. However, if the original block length of 100 characters is retained, there is still a gain of almost 7% in the efficiency without changing the buffer to the optimum block length. To decide when it is practical to adjust the buffer size to the optimum requires knowledge of buffer costs and line rental costs to determine whether the buffer costs offset the line rental savings as a result of increased transmission efficiency.

Conclusions

The importance of the optimum-block-length concept increases with an increase in any of the following parameters: 1. data transmission rate, 2. distance of transmission, and 3. probability of error. When these parameters are all low, as in the top curve ($\delta/\alpha=3$, $p_e=0.00001$) in Fig. 7, the efficiency drops only a few per cent over a range of 50 to 5,000 characters per block. As any one of these parameters increases, the curve of efficiency versus block length sharpens.

For a long-distance transmission, the increase in optimum block length indicates the need for larger buffer storage. The efficiency equation can be used in balancing the cost of the larger buffer required for use of optimum block length against possible reduced line rentals due to a higher effective data transmission rate. The calculation of optimum block length and efficiency permits a comparison of the added rental cost of the local loops to obtain complete 4-wire service with the savings from reduced size of buffer storage.

Table I. Operation at Nonoptimum Block Lengths

			Optimum		Sample Nonoptimum	
f, Bps	p_e	δ/α	$\bar{n}$	E_{max}, Per Cent	n_{no}	E_{no}, Per Cent
1,000	0.00001	24	1,300	83	100	69.1
1,000	0.0001	24	460	77.5	100	68.4
1,000	0.001	24	130	63	100	62.5

In a system where automatic communication between a computer and remote stations requires a periodic feedback signal to protect against unusual events (such as an increase in multiple errors or line failure when there are no spare channels available to the automatic switching system), the details of the situation must be examined to determine whether feedback correction of errors or error-correction by redundancy is more efficient. For some combinations of delay time (δ/α) and probability of a character being in error (p_e), the 7-bit code is more efficient, while for other combinations the 10-bit code is more efficient.

Nomenclature

a = number of bits per character
$E(X)$ = expectation of event X or the average value of X
$E_f(h)$ = transmission efficiency with h redundant bits per character
$\bar{E}_f(h)$ = maximum transmission efficiency obtainable when optimum block length is used
f = transmission rate in bps
$F(p_e,\delta/\alpha)$ = correction factor to obtain exact optimum block length from the approximate equation
h = number of redundant bits per character
i = summation index indicating number of bits or characters in error

k = summation index indicating number of blocks
l = length of transmission channel
m = number of information bits per character
$\bar{n}$ = number of characters per block
n = optimum number of characters per block
$p(x_k)$ = probability of occurrence of event x on the kth trial
p_b = probability that a bit is in error
p_c = probability that a characteris in error
p_c'' = probability that a character is in error due to two or more bits in error within one character
$p_f(k)$ = probability that a block of characters has to be repeated k times to obtain a correct transmission
p_s = probability that a message is correct on the first transmission
$P(X)$ = probability of set of events defined by X occurring
q_c = probability that a character is correct
q_s = probability that a message is not correctly received on one transmission
$S(p_c)$ = a factor in the optimum block length equation which is asymptotic to unity for small values of P_c
T = average time to transmit a block of n characters with all single and multicharacter errors corrected
T_0 = time taken to transmit a block of characters with no errors, but with reply time added to acknowledge message
T_1 = time taken to transmit a block of characters without any feedback signal
X = a set of events such as i characters in error
x_k = an event which occurs on the kth trial
α = time interval of one character
δ = reply delay time interval
δ/α = number of characters which could be transmitted during the reply delay time interval

Appendix I. Derivation of Average Number of Transmissions per Block

To obtain the average time per message T from equation 4, a formula for the expectation is derived from the following definition given by Feller.[8]

Let x be a random variable assuming value $x_1, x_2, \ldots$ with corresponding probabilities $p(x_1)p(x_2)$. The mean or expected value of x is defined by

$$E(x)=\sum_{k=1}^{\infty} x_k p(x_k) \tag{19}$$

provided that the sum converges absolutely.

The grouping in blocks of n characters sets integral values on x_k, so that $x_k=k=1, 2, 3, \ldots$ When the probability of a character being in error is independent, the binomial probability distribution applies, so

$$P\,(i \text{ characters in error})=\frac{n!}{i!(n-i)!}\,p_c{}^{0}q_c{}^{n-i} \tag{20}$$

These repeated independent trials constitute a series of Bernoulli trials as defined by Feller.[9]

Defining $p_f(k)$ as the probability that the block has to be transmitted k times, to obtain a correct transmission on the kth trial yields

$$p_f(1)=q_c{}^n=(1-p_c)^n$$

$$p_f(2)=\left[\sum_{i=1}^{n}\binom{n}{i}p_c{}^i q_c{}^{n-1}\right]q_c{}^n=(1-q_c{}^n)q_c{}^n$$

$$p_f(3)=(1-q_c{}^n)^2 q_c{}^n \tag{21}$$

Extending equations 21 to the general form and substituting it into equation 19 gives the expectation

$$E(X)=q_c{}^n\sum_{k=1}^{\infty} k(1-q_c{}^n)^{k-1} \tag{22}$$

The foregoing gives the average transmission time when substituted into equation 4.

Appendix II. Derivation of Optimum Block Length

Taking the derivative of equation 13 and setting equal to zero gives

$$\frac{dE}{dn}=\left(1+\frac{\delta/\alpha}{n}\right)^{-1}\frac{d}{dn}(1-p_c)^n+(1-p_c)^n\times \left(1+\frac{\delta/\alpha}{n}\right)^{-2}\left(\frac{\delta}{\alpha}\right)n^{-2}=0 \tag{23}$$

$$\frac{d}{dn}(1-p_c)^n=(1-p_c)^n \ln(1-p_c)$$

Let

$$S(p_c)=[-\ln(1-p_c)]/p_c=1+\frac{p_c}{2}+\frac{p_c{}^2}{3}+\ldots+\frac{p_c{}^{x-1}}{x}+\ldots \tag{24}$$

$$(\bar{n})^2+(\bar{n})(\delta/\alpha)-[(\delta/\alpha)\quad/\ln(1-p_c)]=0$$

$$\bar{n}=\frac{-(\delta/\alpha)\pm\sqrt{(\delta/\alpha)^2+4\delta/\alpha p_c S(p_c)}}{2} \tag{25}$$

Taking the positive root in equation 25 gives the optimum block length

$$\bar{n}=\sqrt{\delta/\alpha p_c}\,F(p_c,\delta/\alpha) \tag{14}$$

where $F(p_c, \delta/\alpha)$ is given by equation 15.

Appendix III. Sample Calculation and Graphical Determination of n̄ and E

Conversion of Bit Error Probability to Character Error Probability

When p_b is the probability that a bit is in error in a binary symmetric channel where bit errors are independent, and it is assumed that a block check character provides a longitudinal check on multiple bit errors, the probability of a character being in error (and being detected at the receiver) is

$$p_c=\sum_{i=1}^{7}\binom{7}{i}p_b{}^i(1-p_b)^{7-i} \tag{26}$$

Equation 26 applies for $m=6$, $h=1$ (single-error-detection code, double and higher bit errors to be detected by longitudinal block check).

Sample calculation for $p_b=0.00014$ is as follows:

$$\begin{aligned}p_c&=7p_b(1-p_b)^6+21p_b{}^2(1-p_b)^5+\ldots\\&=7(0.00014)(0.99986)^6+21(0.00014)^2\times(0.99986)^5+\ldots\\&=0.000\,98+0.000\,00041+\ldots\sim10^{-3}\end{aligned}$$

For $m=6$, $h=4$ (single-error correction, double and higher order bit error to be detected by longitudinal block check), the corresponding formula is

$$p_c''=\sum_{i=2}^{10}\binom{10}{i}p_b{}^i(1-p_b{}^i)^{10-i}$$

$$\begin{aligned}&=p_c''=45p_b{}^2(1-p_b)^8+120p_b{}^3(1-p_b)^7+\ldots\\&\simeq0.000\,000\,9+0.000\,000\,000\,336+\\&\simeq0.000\,000\,9\simeq10^{-6}\end{aligned} \tag{27}$$

$$p_c''\sim p_c{}^2,\ \text{for } p_c<0.001 \tag{28}$$

Graphical Determination of n̄ and E

Condition serial transmission, 1,000 bps, $p_c=0.0001$, $l=3{,}000$ miles.

Step 1:

$$\alpha=\frac{a}{f}=\frac{7\text{ bits per character}}{1{,}000\text{ bps}}=0.007\text{ second per character}$$

Step 2: Take $l=3{,}000$ miles, assume typical transmission line with a total of 0.300 (second) delay time for 2-way propagation and resetting of the echo suppressors. Assume that two characters are used for block-check and answer-back signals and that the logical operations can be completed within one character time, so that $\delta/\alpha=3+(0.300/0.007)=47$.

Step 3: In Fig. 2 the intersection of the curve $\delta/\alpha=47$ with $p_c=0.0001$ is marked as point A. This gives optimum block length $\bar{n}=640$, and maximum efficiency $\bar{E}=(6/7)86=74\%$.

Step 4: To determine the efficiency when 100 characters are used as the block length so that $n=100$ characters (see Fig. 4), the exact result can be obtained from equation 13 or the efficiency can be obtained by interpolating between the curves of Fig. 4. Point A for $\bar{n}=640$ and point B for $n=100$ are shown in Fig. 4. From point B it is determined the the efficiency is reduced to 57% when a block length of 100 characters is used.

References

1. Error Checking for 5-Channel Telegraphic Tape, **R. A. Barbeau.** *AIEE Transactions*, vol. 77, pt. I, May 1958, pp. 190–93.
2. A Mathematical Model for Determining the Probability of Undetected Errors in Magnetic Tape Systems, **M. Schatzoff, W. B. Harding.** *Journal of Research and Development*, International Business Machines Corporation, New York, N. Y., vol. 1, Apr. 1957, pp. 177–80.
3. Message Redundancy Versus Feedback for Reducing Message Uncertainty, **W. B. Bishop, B. L. Buchanan.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., vol. 5, pt. 2, 1957, pp. 33–39.

4. **Protection of Service in the TD-2 Radio Relay System by Automatic Switching, I. Welber, H. W. Evans, G. A. Pullis.** *Bell System Technical Journal*, New York, N. Y., vol. 34, May 1955, pp. 473–510.

5. **An Introduction to Probability Theory and Its Applications, Vol. 1 (book), William Feller.** John Wiley & Sons, Inc., New York, N. Y., second edition, 1957, p. 156.

6. **Error Detecting and Error Correcting Codes, R. W. Hamming.** *Bell System Technical Journal*, vol. 20, Apr. 1950, pp. 147–60.

7. **Principles of Electricity Applied to Telephone and Telegraph Work** (book). Long Lines Department, American Telephone and Telegraph Company, New York, N Y., 1953 edition, pp. 274–76.

8. **Feller,** *op. cit.*, p. 207.

9. *Ibid.*, pp. 135, 137.

Volt-Second Transfer Efficiency in Fast-Response Magnetic Amplifiers

Part I: N^2/R and Control

T. J. PULA
ASSOCIATE MEMBER AIEE

TO MAKE magnetic amplifier theory mathematically tractable, numerous analyses which have appeared in the literature employ assumptions leading to either "ampere-turn control" or "volt-second control" idealizations. While either set of assumptions leads to quite good prediction of operation for simple saturable reactors, both simplified approaches break down at several important points when extended to self-saturating magnetic amplifiers. The regions of inadequate correlation have been explored extensively in recent years in terms of ampere-turn control, especially for self-saturating amplifiers exhibiting simple time delay. However, the limitations of the more recent volt-second control approach[1,2] have received relatively little attention, especially for amplifiers exhibiting only dead-time (i.e., so-called "fast-response" amplifiers).[3–5]

The object of this paper is to extend the volt-second control approach into the region of finite signal resistance and non-ideal flux-current loops for fast-response amplifiers. In the process, certain discussions of doublet amplifiers in terms of ampere-turn control will be pursued for finite signal resistance and an attempt made to clarify the complementary relationships existing between these approaches for realistic flux-current loops. The term "doublet" is used here to describe any amplifier with a configuration of control windings which results in an exponential change in output for a step change in input.

As a result of a detailed study of the qualitative mechanics of operation of self-saturating magnetic amplifiers,[6] it has been shown that several different modes of operation are imposed upon a core material in the various types of self-saturating circuits. It is therefore quite difficult to devise one set of measurements to specify completely the significant core material parameters for these disparate circuit types. To clarify this point, each of these types will now be reviewed briefly with more specific reference to the role of the magnetic core material.

Ramey-type circuits are based on the determinate volt-second control approach and, when the required condition is met, namely $R_s \cong 0$, the operation of the amplifier is insensitive to core characteristics. However, this condition is not reasonable for a first-stage amplifier; therefore the usefulness of Ramey-type circuits is confined to second-stage amplifiers. In this application they are quite useful, especially when angle volt-second control is provided by a switching impedance in the first stage. Therefore no sophisticated core gain measurements are necessary for applications where the assumptions imposed by this type of control can be met.

Doublet circuits employ the fed back volt-seconds from the gating source to aid reset, thus increasing the obtainable volts-per-volt gain at the sacrifice of dynamic response. Their dependence on core characteristics will be discussed later with the aid of the fundamental parameter N^2/R and of incremental volts per ampere-turn gain. N^2/R is defined as the summation of the quotients of the square of the number of turns and the circuit resistance for all control circuits. As in the qualitative descriptions of reference 6, the bridge circuit will be investigated as representative of the class.

In designing a 2-stage half-wave amplifier which obtains its signal from a source with appreciable impedance, a problem common to various types of modern fast-response amplifiers presents itself. The second stage of these amplifiers, designated here as hybrid circuits, requires no close matching of core material gain characteristics, once the type of material has been properly chosen. The first stage, however, presents a problem in volt-second transfer efficiency because of the finite nature of R_s. As an application of the ideas developed in this paper, the optimization of the first stage of one amplifier in this class will be the subject of a companion paper.

Now, if ideal core material were assumed in such a first stage, determinate volt-second control would be obtained because no current would be required to flow through the reactor to accompany its change of flux, i.e., the net ampere-turns to the reactor would be zero while it was in an unsaturated state. However, since such a material does not exist, the amount of voltage which will appear at the terminals of the reactor cannot be predicted as soon as the voltage which is supplied from the signal source is known. Thus, an extra degree of freedom has been introduced because the difference between the reactor voltage and the signal voltage is a function of the resistor voltage, which in turn is equal to the signal current times the signal circuit resistance. Such a control regime has already been designated in reference 6 as "indeterminate volt-second control," wherein there is an interrelationship between the current flowing in the signal circuit and the flux reset produced. It is specifically in this area where experimental measurements of magnetic

---

Paper 58-1193, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted April 2, 1958; made available for printing August 11, 1958.

T. J. Pula is with Westinghouse Electric Corporation, Baltimore, Md.

The author wishes to thank Profs. T. Larsen and C. F. Miller of the Electrical Engineering Department, The Johns Hopkins University, for guidance during this work. Stimulating discussions with G. E. Lynn, J. F. Ringelman, and F. G. Timmel of Westinghouse Air Arm Division, Baltimore, are also gratefully acknowledged. This paper presents a portion of the work submitted to The Johns Hopkins University, Baltimore, Md., in partial fulfillment of the requirements for the degree of Doctor of Engineering.

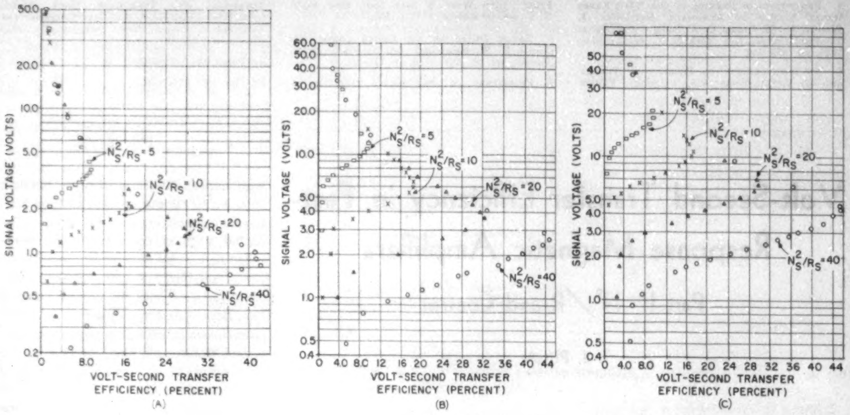


Fig. 1. Volt-second transfer efficiency for 2-mil, Hymu 80, 3/8- by 1/2- by 1/8-inch cores, N_s^2/R_s as parameter. A—$N_s=100$. B—$N_s=300$. C—$N_s=500$

core characteristics are required in the process of the design of self-saturating magnetic amplifiers. This has led some investigators,[7–9] to the extreme of simulating a control circuit current source and obtaining characteristics of the core material with a control circuit resistance approaching infinity. The disadvantage of this approach, however, is that the voltage supplied for control and, consequently the volts-per-volt gain associated with the reset obtained by this voltage, become obscure. In addition, the materials are not used with signal circuits satisfying the conditions imposed by the core measurement configurations. These considerations have led the author to pursue the investigation of indeterminate volt-second control in terms of measurements aimed at obtaining the volt-second transfer efficiency characteristics of reactors for the regime where the signal circuit resistance is finite.

Volt-Second Transfer Efficiency Measurements

The Logan (or half-wave) circuit with square-wave excitation, but without bias was chosen for these measurements. This combination is considered to be the most basic configuration: the Logan circuit because it represents the classic single-core self-saturating circuit, the square-wave excitation to minimize "back-firing," and the elimination of bias in the initial measurements to avoid coupling complications between control circuits.

To facilitate eventual correlation of these measurements with existing core evaluation technology, individual circuits from the constant-current flux-reset tester were used with minimal modifications. This test method has been thoroughly described in the literature.[7–9] The basic change made in the test method was the elimination of the inductor in each control circuit. This inductor is used in the constant-current test (along with low N and high R) to approximate $N^2/R=0$, whereas one object of the present study is to investigate the significance of N^2/R as a parameter.

The measurements consisted of recording the average voltage transferred to the reactor for various voltages applied by the signal source. These measurements, for various signal turns and signal circuit resistance values for a 2-mil Hymu *80* material, are shown in Fig. 1.

It may be seen in Fig. 1 that, for equal N^2/R values, the peak volt-second transfer efficiency values are equal but that they occur at a voltage which is higher in proportion to the number of signal turns used. This means that the accompanying average ampere-turn values are equal. When the results for equal N^2/R values are plotted not as volt-second transfer versus signal voltage but as volt-second transfer versus volts per turn applied, as shown in Fig. 2, it is found that the complete curves for equal N^2/R values coincide with each other within the accuracy of the experimental method. Similar results have been obtained for a 2-mil *50-50* NiFe material but are omitted for brevity. Therefore, even for the case of square-loop materials, where the concept of linear inductance breaks down, N^2/R remains a basic reactor parameter, and plots of volt-second transfer efficiency versus volts per turn with N^2/R as a parameter provide a basic set of core material reset characteristics.

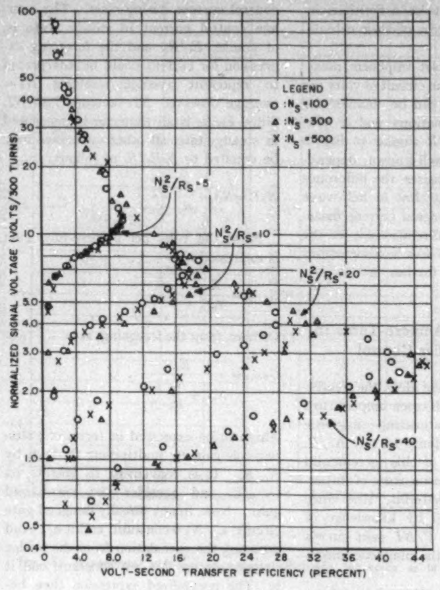
It is interesting to note from this normalization of the signal voltage that E_s/N then has the units of flux change with respect to the average core voltage (from

$\Delta\phi = \mathrm{bu}/\int e\,dt$)

It appears, then, that the obtained change of flux for each particular E_s is a constant percentage of the impressed change of flux for various combinations of N_s and R_s as long as the N^2/R values are equal. It is only when R_s approaches zero or the flux-current loop becomes ideal that a flux change is obtained which corresponds to the value which would be calculated from the normalized applied voltage.

Further Investigation of the Parameter N²/R

The emergence of N^2/R as a fundamental parameter in the volt-second transfer efficiency characteristics of the Logan circuit during reset allows a very interesting observation concerning voltage versus current in control of any magnetic amplifier. From Faraday's Law:



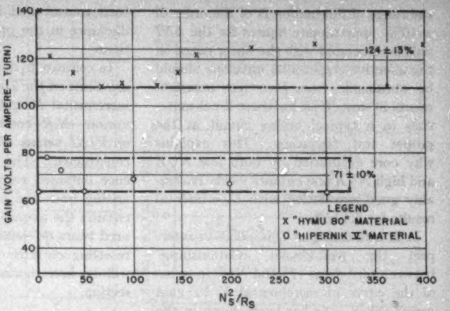


Fig. 2. (left). Normalized volt-second transfer efficiency for cores as in Fig. 1

Fig. 3 (above). Volts per ampere-turn gain versus N_s^2/R_s in bridge circuit for 2-mil, Hymu 80 and Hipernik V cores

$$e=N\frac{d\phi}{dt}$$

$$=AN\frac{d(\mu H)}{dt}$$

$$=AN\,\frac{d(\mu kNi/l)}{dt}$$

$$=\frac{kA}{l}\,N^2\,\frac{d(\mu i)}{dt}$$

$$e=\frac{kA}{l}\,N^2\mu\,\frac{di}{dt}+\frac{kA}{l}\,N^2 i\,\frac{d\mu}{dt}$$

In this equation, even though the terms $\mu\ di/dt$ and $i\ d\mu/dt$ cannot be evaluated in general, the coefficient of each of the terms equated to e contains the factor N^2. This, of course, is what would be expected from the transformer approach. However, it is only when each term of this expression is divided by $R\neq 0$ that N^2/R appears in the coefficient. And in this case it is related to e/R, a core-current expression. Therefore, as soon as the determinate volt-second control idealization is abandoned, the emergence of N^2/R as a basic parameter equivalently expresses the fact that the current accompanying reset also becomes basic.

In view of the meaning of N^2/R in a doublet amplifier, although it is true that in steady state $I_s=E_s/R_s$, the volt-seconds induced into the signal circuit during each reactor's prefiring interval may be considered to contribute to reset. In fact, for high N^2/R the induced volt-second area on any reset half-cycle greatly exceeds the volt-seconds providedby $E_s\times \Delta t$. This fact has led to many attempts to relate the incremental ampere-turns associated with the carrier frequency sinusoidal voltage loop to the incremental resetting ampere-turns required for control. Care must be exercised in this regard, however, because a large number of half-cycles is required to obtain the new steady-state condition and the basic function of E_s is gradually to shift the steady-state operating point of the transient ratcheting process, as discussed in reference 6. Thus, on individual reset half-cycles during the transient, the maximum gating ampere-turn value varies for firing angles greater than 90 degrees, and an altered domain configuration at the beginning of reset should be expected. Therefore the incremental resetting ampere-turns cannot be associated in a straightforward way with the carrier frequency sinusoidal voltage loop for large changes of output. For this reason it is helpful to visualize an effective trapezoidal envelope similar to that developed in the discussion of the Logan circuit in reference 6.

In contrast, for low N^2/R, the volt-seconds induced into the control circuit have a negligible effect. Reset is produced by control alone. Assuming for definiteness that the bridge circuit is biased to cutoff, the ΔNI product required from the signal source is that value corresponding to the maximum flux excursion of the effective trapezoidal envelope. It should be noted for this case that this ΔNI product cannot be associated with any single flux current loop.

To produce the required change in flux for the low N^2/R case, a corresponding average $d\phi/dt$ must be obtained so that the required $\Delta\phi$ is obtained in 1/2 cycle of the carrier frequency. This is the same average $d\phi/dt$ represented by the carrier frequency voltage source loop and might lead to a prediction of comparable ΔNI products. However, the losses associated with the flux change (and therefore the ΔNI required) are a function of the instantaneous $d\phi/dt$ rather than the average $d\phi/dt$. Therefore, it cannot be stated a priori that the V/NI_{avg} gain will be strictly independent of N^2/R. However, there is no straightforward relationship showing variation of V/NI gain with N^2/R.

Measurements of incremental volts per ampere-turn gain versus N^2/R for bridge circuits constructed from Hymu *80* and Hipernik *V* materials were obtained and the incremental volts per ampere-turn gain versus N^2/R plotted in Fig. 3. Although some functional relationship exists in these curves, there is no large magnitude trend. Furthermore, since the range of fluctuation is of the order of ±10%, approximate figures for the ΔNI values associated with the linear range on the effective trapezoidal envelope should be obtainable from core test measurements at any N^2/R for these 2-mil materials in a typical bridge circuit at the proper test frequency. This explains why core evaluation by both low N^2/R and high N^2/R test circuits yields reasonably good correlation with experimental results for doublets.[7–10]

It is interesting at this point to interpret the well-known relationship,[11] between time delay (T) and N^2/R in terms of the curve of incremental V/NI gain versus N^2/R. As has been shown in the plot of Fig. 3, the ΔNI required for full ΔE_L over the linear range is, to a good approximation, independent of N^2/R. Now, since in steady state, $\Delta E_s = \Delta I_s R_s$, the $\Delta E_L/\Delta E_s$ ratio reflects the $1/R_s$ $(\Delta E_L/\Delta I_s)$ ratio. (This is true because, although R_s can be any finite value in this discussion, it is held fixed at one setting while any particular $\Delta E_L/E_s$ curve is taken.) Now, this $\Delta E_L/\Delta I_s$ ratio is finite only if the left flank of the effective trapezoidal envelope is not perfectly vertical. Correspondingly, therefore, the Harder and Horton expression for time delay

$$T = \frac{1}{2f_0}\,\frac{\Delta E_L}{\Delta E_s}\,\frac{N_s}{N_L}$$

f_0 = supply frequency

which may now be written as

$$T = \frac{1}{2f_0}\,\frac{N_s}{N_L}\,\frac{1}{R_s}\,\frac{\Delta E_L}{\Delta I_s}$$

or

$$T = k\,\frac{\Delta E_L}{\Delta I_s}$$

for any particular amplifier, yields a finite value for the time constant only if the left flank of the effective trapezoidal envelope is nonvertical.

The central point to be emphasized here is that this ΔI_s, associated with the change of E_L requires a considerable number of half-cycles of the supply frequency to be realized. Thus, although concepts of incremental volt-second transfer are applicable in each reset half-cycle, control is produced only by a series of such increments in conjunction with gradually shifting boundary conditions, i.e., an integrated volt-second transfer mechanism. In the opinion of the author, it was largely this integration process and the associated constancy of the $\Delta N_s I_s$ which led many early writers to discuss doublet amplifiers in terms of ampere-turn control and invoke a fictitious inductance in the process of their calculations.

In contrast to doublet amplifiers, measurements on the Logan circuit of volts per incremental NI gain can be obtained for various N,R combinations and a plot of V/NI versus N^2/R similar to Fig. 3 constructed. The well-known dependence obtained emphasizes the difference arising from the fact that in half-wave circuits the average signal current measured bears no general relationship to the resetting current. These considerations will be investigated further in the next section.

Volt-Seconds and Ampere-Turns in Magnetic Amplifier Control

If core materials were ideal, the steady-state incremental volts per ampere-turn gain of any self-saturating magnetic amplifier would be infinite for any N^2/R. Therefore the value of this incremental V/NI gain provides a measure of imperfection in the core material. And since designs are optimized by knowledge of imperfections, these V/NI gain curves form a basic link in the magnetic amplifier design art.

It is interesting to attempt to extend the idea of constant incremental ampere-turns to a fast-response indeterminate volt-second controlled stage where the average incremental NI measured in the signal circuit cannot be directly associated with reset, since reset is obtained only on alternate half-cycles. (In such cases the induced current from E_{ac} during the prefiring portion of a gating half-cycle does not produce any core reset as in the bridge circuit.)

For this purpose, consider the governing equations of the Logan circuit during a reset half-cycle. From Kirchhoff's voltage law

$$E_c = i_c R_c + e_{\text{core}}$$

Now, by basic magnetic-amplifier design, for complete reset

$$\bar{e}_{\text{core}} = \frac{N_c}{N_L}\,\bar{e}_{ac}$$

Therefore, in terms of average values for complete reset

$$\bar{\imath}_c = \frac{E_c - (N_c/N_L)\bar{e}_{ac}}{R_c}$$

Multiplying through by N_c yields

$$N_c\bar{\imath}_c = \frac{N_cE_c - (N_c^2/N_L)\bar{e}_{ac}}{R_c}$$

Note carefully here that $\bar{\imath}_c$ is not the average control current but the average control current during reset. The average control current in steady state is, of course, E_c/R_c, and the foregoing expression for current could be interpreted to represent average resetting NI = average observed NI − average loss NI. Since E_c/R_c is all that can be measured in steady state, all other quantities may be equated to N_cE_c/R_c as follows:

$$N_cI_c = N_c\bar{\imath}_c + \frac{N_c^2}{R_c}\,\frac{\bar{e}_{ac}}{N_L}$$

Now, the volts per ampere-turn gain is expressed as

$$G_{v/NI} = \frac{E_0}{N_cI_c}$$

Or here, from the foregoing, as

$$G_{v/NI} = \frac{E_0}{N_c\bar{\imath}_c + \dfrac{N_c^2}{R_c}\,\dfrac{\bar{e}_{ac}}{N_L}}$$

This can be expressed in terms of volts-per-volt gain by multiplying through by N_c/R_c. It is convenient to operate on $G_{v/v}/E_0$ and consider the normalized gain. Now, in any already designed gate circuit, $\bar{e}_{ac}/N_L$ = constant, call it a. And pursuing the speculation from the bridge circuit, assume $N_c\bar{\imath}_c$ = a constant, call it b. The normalized expression then becomes

$$\frac{G_{v/v}}{E_0} = \frac{N_c}{bR_c + aN_c^2}$$

Taking the partial derivative of this expression with respect to N_c for R_c fixed $\neq 0$, equating to zero, and solving for N_c^2/R_c

$$\frac{N_c^2}{R_c} = b/a$$

This then represents the value of N_c^2/R_c which will maximize the volts-per-volt gain. Substituting back this value for N_c^2/R_c yields

$$G_{v/NI} = E_0/2b$$

Therefore, for maximum volts-per-volt gain with R_c fixed $\neq 0$, a value of N^2/R which makes the V/NI gain equal to 1/2 $(G_{v/NI})_0$ should be chosen.

In terms of incremental volt-second transfer efficiency, therefore, the N_c^2/R_c which gives $G_{v/NI} = 1/2$ $(G_{v/NI})_0$ gives maximum $G_{v/v}$ when N is varied for fixed $R_c \neq 0$. But this result is obtained by assuming the product (N_ci_c) equal to a constant for a given material configuration and frequency for full linear ΔE_0. Therefore, for maximum $G_{v/v}$ for fixed R_c this corresponds to $I_c = 2\bar{\imath}_c$. And in steady state, $I_c = E_c/R_c$. Therefore $i_cR_c = E_c/2$. In other words, maxi-

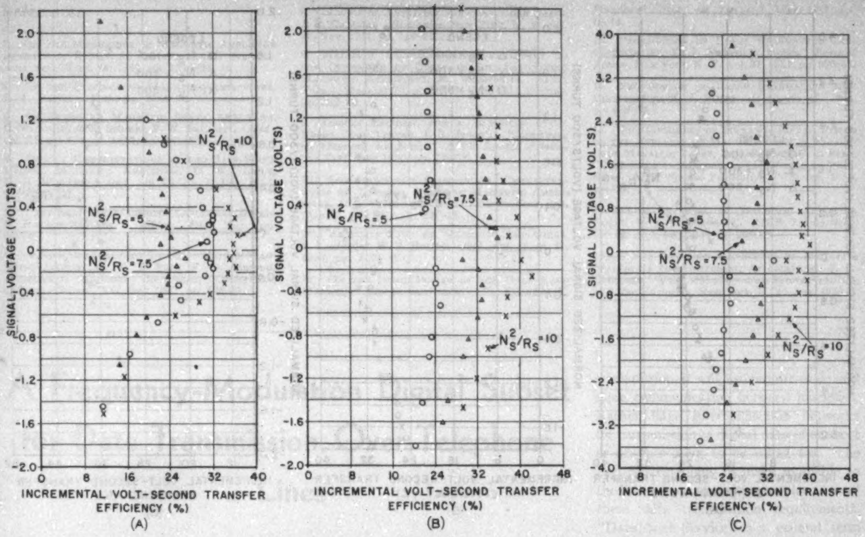


Fig. 4. Incremental volt-second transfer efficiency for cores as in Fig. 1, with bias. A—N_c=100. B—N_c=300. C—N_c=500

mum $G_{v/v}$ is obtained for fixed $R_c \neq 0$ when, in terms of average values, half the voltage during reset is allowed to appear across R_c. But this is just what is obtained with an incremental volt-second transfer efficiency of 50%. This result has been obtained recently[4,12] without explicit attention to average values.

It is very interesting to attack this same derivation by applying directly the observation for doublets that the average $(d\phi/dt)$ must be equal for all N^2/R values. The governing equations for the Logan circuit during reset can be manipulated as follows:

$$E_c = i_c R_c + N_c \frac{d\phi}{dt}$$

Multiplying through by N_c/R_c and rearranging to isolate the resetting mmf (magnetomotive force) yields

$$N_c i_c = \frac{N_c}{R_c} E_c - (N_c^2/R_c)\frac{d\phi}{dt}$$

Now, consider the value of N_c which will maximize this mmf for complete reset. For constant $R_c \neq 0$, by partial differentiation with respect to N_c and equating to zero,

$$\frac{E_c}{R_c} - \frac{2N_c}{R_c}\frac{d\phi}{dt} - \frac{N_c^2}{R_c}\frac{\partial(d\phi/dt)}{\partial N_c} = 0$$

Considering this expression in terms of average values and tentatively extending the conclusion from doublets, i.e., $\overline{(d\phi/dt)}$ = constant, yields

$$E_c = 2N_c\left(\overline{\frac{d\phi}{dt}}\right)$$

Therefore, for constant $R_c \neq 0$, the average net mmf during reset is maximized by choosing N_c so that

$$\frac{N_c\left(\overline{\frac{d\phi}{dt}}\right)}{E_c} = 0.50$$

But this is equivalent to obtaining a volt-second transfer ratio (VSTR) of 50%, since

$$\text{VSTR} = \frac{\int_0^{T/2} e_{\text{core}}\,dt}{E_c \times T/2}$$

$$\text{VSTR} = \bar{e}_{\text{core}} \times \frac{1}{E_c}$$

and

$$\bar{e}_{\text{core}} \equiv N_c\left(\overline{\frac{d\phi}{dt}}\right)$$

Therefore, both the ampere-turn control approach assumption of $N_c\bar{i}_c$ = constant for complete reset and the volt-second control approach assumption of $\overline{(d\phi/dt)}$ = constant for complete reset both yield the same result of adjusting N_c, for fixed $R_c \neq 0$, to obtain a volt-second transfer ratio of 50%.

Apart from the usefulness of this conclusion in optimizing the volts-per-volt gain of certain types of self-saturating magnetic amplifiers, the correspondence of both ampere-turn and volt-second approaches sheds considerable light on the fundamental physical processes in control. In the first place, analyses using practical core materials cannot neglect the incremental ampere-turns required to produce a change of flux level and, contrariwise, the choice of parameters which maximizes the average resetting magnetomotive force at the same time produces the optimum magnitude of this $\overline{(d\phi/dt)}$, and consequently the maximum reset with the associated maximum volts-per-volt gain.

These complementary assumptions (i.e., $N\bar{i}_c$ = constant and $\overline{(d\phi/dt)}$ = constant) are quite reasonable in terms of modern domain theory because they imply that for a given average incremental H field (proportional to $\Delta N\bar{i}_c$), the same average domain wall velocity, proportional to $\overline{(d\phi/dt)}$, is obtained regardless of how this configuration was produced in the circuit. And it has been shown[12-16]

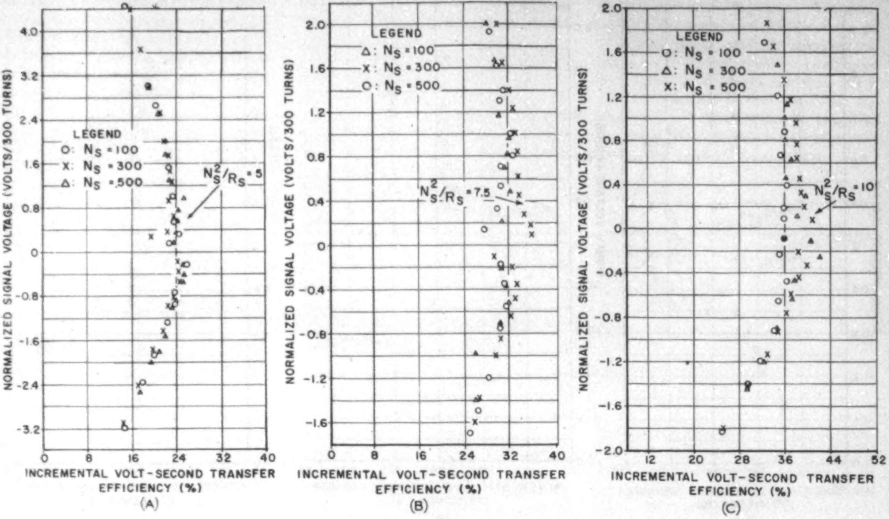


Fig. 5. Normalized incremental volt-second transfer efficiency for cores as in Fig. 1, with bias

that this proportionality constitutes one of the basic relationships in domain wall dynamics. The significance of these conclusions in terms of incremental volt-second transfer efficiency measurements will be investigated in the following section.

Incremental Volt-Second Transfer Efficiency Measurements

In the volt-second transfer efficiency curves of Fig. 2 the volt-second transfer efficiency versus volts per turn applied is quite variable and, as would be expected, this produces a quite nonlinear volts-per-volt amplifier gain curve. For this reason, it is general design practice to provide some sort of bias in order to shift the operating point of the magnetic reactor to a more nearly linear portion. In terms of this approach, therefore, information concerning the incremental volt-second transfer efficiency of resetting reactors must be obtained. This can be done by biasing the Logan circuit to half of maximum volt-second transfer, with a separate winding having a negligible N^2/R.

For the measurements obtained with the special test circuit used, the N^2/R of the bias circuit was adjusted to approximately 0.30. Later measurements will show that the fundamental nature of N^2/R as a parameter is not jeopardized by this procedure, nor is the character of volts per turn as a normalized signal. When measurements of incremental volt-second transfer efficiency versus signal voltage are plotted, as in Fig. 4, the curves are seen to be symmetrical and considerably more vertical than without bias. Furthermore, when these curves are replotted versus volts per turn applied as in Fig. 5, the curves for a particular N^2/R show strong correlation, i.e., the incremental volt-second transfer efficiency remains substantially constant over a considerable swing of the input signal and an incremental efficiency can be associated with each N^2/R. This material characteristic is really the source of the linear portion of the volts per volt or volts per incremental ampere-turn gain curve of self-saturating magnetic amplifiers.

Now, as a series of curves of incremental volt-second transfer efficiency versus volts per turn applied is obtained, it is seen that the curves show a shift in percentage with N^2/R. A fundamental characteristic of a particular core configuration at a given frequency, therefore, is the N^2/R value corresponding to a particular incremental volt-second transfer efficiency percentage. In a companion paper tentatively titled "Part II: N^2/R as a Design Parameter," it will be shown that for a very useful type of fast-response amplifier the volts-per-volt gain can be optimized by choosing an effective incremental volt-second transfer efficiency of 50% for a fixed value of control circuit resistance other than zero. In other words, in a particular design for such an amplifier, when signal resistance is fixed at some allowable minimum, these plots can be employed to pick the optimum value of N_s, i.e., to pick the value of signal turns to yield 50% effective incremental volt-second transfer efficiency for that resistance and, consequently, to yield the maximum volts-per-volt gain in the amplifier.

Conclusions

The volt-second control approach has been extended to consider a finite control circuit resistance and nonideal core characteristics by obtaining volt-second transfer efficiency characteristics for reactors composed of square-loop magnetic materials. The nature of N^2/R as a basic reactor parameter has been demonstrated experimentally and this knowledge used in deriving expressions which indicate the complementary relationship existing between volt-seconds and ampere-turns in magnetic amplifier control.

References

1. On the Mechanics of Magnetic Amplifier Operation, **Robert A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. On the Control of Magnetic Amplifiers, **R. A. Ramey.** *Ibid.*, pp. 2124–28.

3. An Improved Magnetic Servo Amplifier, **C. W. Lufcy, A. E. Schmid, P. W. Barnhart.** *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 281–89.

4. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, **C. W. Lufcy, H. H. Woodson.** *Ibid.*, vol. 73, pt. I, July 1954, pp. 220–26.

5. Fast-Response Magnetic Amplifiers, **Gerald E. Hughes, Harvey A. Miller.** *Ibid.*, Mar., pp. 69–75.

6. A New Approach to Self-Saturating Magnetic Amplifiers, **T. J. Pula, F. G. Timmel.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XII, Apr. 1957, pp. 425–43.

7. Magnetic Amplifier Gapless Core Tests, **J. R. Conrath.** *Electronics*, New York, N. Y., vol. 25, Nov. 1952, pp. 119–21.

8. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, **R. W. Roberts.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

9. Design Criteria for a Practical Flux-Reset Core Tester, **J. R. Jaquet.** "Conference on Magnetic Amplifiers," *AIEE Special Publication T-86*, July 1956, pp. 74–105.

10. Low-Impedance Operational Characteristics of Toroidal Cores in Magnetic Amplifiers, **C. L. Boyajian.** *AIEE Transactions*, vol. 75, pt. I, July 1956, pp. 307–13.

11. Response Time of Magnetic Amplifiers, **E. L. Harder, W. F. Horton.** *Ibid.*, vol. 69, pt. II, 1950, pp. 1130–41.

12. On the Input Power of the Half-Wave Magnetic Amplifier Circuit, **T. Kikuchi, K. Murakami.** *Ibid.*, vol. 76, pt. I, Mar. 1957, pp. 10–14.

13. Propagation of Large Barkhausen Discontinuities, **K. J. Sixtus, L. Tonks.** *Physical Review*, New York, N. Y., vol. 37, 1931, pp. 930–59.

14. Propagation of Large Barkhausen Discontinuities, **K. J. Sixtus.** *Ibid.*, vol. 48, 1935, pp. 425–30.

15. The Influence of Pulsed Magnetic Fields on the Reversal of Magnetization in Square-Loop Metallic Tapes, **D. S. Rodbell, C. P. Bean.** "Conference on Magnetism and Magnetic Materials," *AIEE Special Publication T-78*, Oct. 1955, pp. 219–27.

16. Flux Reversal in Magnetic Amplifier Cores, **F. J. Friedlaender.** *AIEE Transactions*, vol. 75, pt. I, July 1956, pp. 268–78.

17. Pulse Response Properties of Magnetic Materials for Switching Applications, **J. D. Childress.** *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., Aug. 1956,

A Frequency-Modulation Digital Subset for Data Transmission Over Telephone Lines

LAURANCE A. WEBER
MEMBER AIEE

Synopsis: Modern business is embracing automation. In the business field, automation is often termed integrated data processing (IDP). This involves the use of large-scale digital computers as fundamental elements in the bookkeeping and routine activities of business. The communication system role in IDP is to transmit digital data from one location to another. This may involve sending data from subsidiary centers to a processing center or it may involve transmission of data from the central location to other business offices. The telephone companies are especially well fitted to handle the needs of business customers in this field because of their well-developed wire line distribution networks. The Bell System expects to offer digital transmission services over normal telephone lines. This is the Bell System Dataphone Service.

As one of a line of instrumentalities for the Dataphone service, an FM (frequency-modulation) digital subset has been developed for transmitting sequential binary data over telephone lines. This equipment accepts sequential binary data from customers' machines at rates up to about 1,000 bps (bits per second), modulates this base band signal to FM signals appropriate to the bandwidth available on telephone channels, transmits this signal to the line, detects the signal at the receiving end, and reconverts the line signals to d-c serialized binary data of the same form as originally presented.

This paper indicates some of the design problems for such transmitting-receiving equipment and also indicates some of the possibilities which appear to be forthcoming in the Dataphone service. At present a limited number of circuits of the type to be described are being used by various Bell System companies to test and assess the features and facilities which will be needed by Bell System customers for the Dataphone service.

The FM modulator-demodulator employed in the digital subset is fully transistorized. It represents a first-order effort at the solution of the transmission problem of sending data over switched telephone connections. Since the binary information rate (up to about 1,000 bps) is comparable to the transmission carrier frequency, the detecting means must be able to reconstruct the digital data even when there are only a few cycles of carrier signal per bit. The demodulator uses the principle of zero-crossing detection to accomplish this. The digital subset has been designed for maximum flexibility in working with existing input-output devices. Flexibility has been included to allow use with equipment still in the preliminary development stage.

The digital subset is arranged so that a Dataphone station may be operated remotely on an unattended basis in conjunction with certain auxiliary apparatus. This will allow pickup and delivery of data to remote stations during nonbusiness hours.

THE DEMAND of modern business for IDP will place a large burden on the communication companies. It is obvious in this age of decentralization that, just as business requires vast amounts of communication for voice transmission, it will also require more and more communication between machines. Rather large interconnection networks for transmission of digital data of one sort or another seem to be called for. The Bell System plans to make available to the business community a service to meet these data transmission requirements. "Dataphone Service" is a general term which describes the transmission of data over telephone voice facilities. The telephone companies are uniquely able to provide service of this type because of the general availability of voice channels and the well-developed methods of interconnecting customers on a demand-real-time basis. At present, it appears that most of the data transmission requirements that businesses have will be met with the use of normal telephone voice channel bandwidths.

This paper will discuss some of the general problems involved in data transmission over telephone voice facilities. It will also describe, as one of the instrumentalities of the Dataphone service, an FM digital subset which has been designed to accept serialized d-c binary input information at bit rates up to about 1,000 bps and to transmit this information at appropriate carrier frequencies over a regular telephone channel. In the future it may turn out that the amount of data which must be sent between two points in an IDP network may

---

Paper 58-1204, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 17, 1958; made available for printing August 14, 1958.

Laurance A. Weber is with Bell Telephone Laboratories, Inc., New York, N. Y.

For valuable contributions in this development the author is greatly indebted to many of his colleagues at the Bell Telephone Laboratories, especially to R. M. Gryb, J. R. Davey, R. O. Rippere, W. N. Toy, R. K. Townley, and T. A. Jones.

involve bit rates higher than this. This will require more efficient utilization of the voice bandwidth by more sophisticated coding or modulation schemes or the application of broader bandwidths to the problem. Some interesting techniques toward this end have been reported.[1,2]

The Data Transmission Problem

The data transmission problem with higher bit rates and the use of voice-band channels is in a sense merely an extrapolation of the familiar telegraph transmission problem which has received much attention over the years. Theory states that the channel capacity C, in the Shannon sense, is given by

$$C=f_0 \log_2\left(1+\frac{S}{N}\right) \quad \text{bps} \qquad (1)$$

where f_0 = the bandwidth in cps (cycles per second) and S/N is the power signal-to-noise ratio. For bandwidths of, say, 2,500 cps and S/N ratios of, say, about 20 db (decibels) or so, the channel capacity works out to be in the order of about 15,000 bps. It is also known that one should be able to signal up to the channel capacity in equation 1 with an arbitrarily small error rate by using suitably selected coding. However, when the communication channel is a telephone line, several factors compel constraint to a practical rate of perhaps a tenth or a little more of the theoretical capacity. In the first place, the amplitude-phase characteristics of the channels differ from that assumed by the theory; delay distortion at the extremes of the band, for example, must be considered. Secondly, noise, particularly of the impulsive type, and other momentary disturbances often cause difficulty which is most severe at the higher bit rates. There are also net loss changes in the circuits. These are all different from the simple Gaussian interference assumed in the derivation of equation 1. Maximum exploitation of the capacity relies on coding techniques which usually result in prohibitive terminal costs, at least for general use. A further problem is introduced when data channels are to be set up on a demand basis, since tandem combinations of all the many differing types of telephone transmission facilities will be encountered. For an excellent discussion of the general problem refer to the paper by Mertz and Mitchell;[3] their paper indicates the possibilities of amplitude modulation vestigial sideband and FM for high-speed data transmission over message facilities. The FM digital subset, to be described in this paper, was built to exploit this method of transmission.

The Digital Subset

The subset makes it possible for business machines or computers to "talk" to each other over telephone lines. The subset is arranged for alternate use of the voice line for normal speech communication and for data. The subset accepts binary digital information from the customer's equipment, converts it to suitable frequency-modulated voice tones for transmission over the channel, recovers the binary data signals at the receiving station, and passes this information reshaped but not retimed to the receiving customer's data equipment.

The basic transmission scheme is simple frequency shift carrier operation. When a binary 1 or "mark" is presented to the subset a voice frequency F_1 is transmitted;

Vol.
Pt. 1
1958

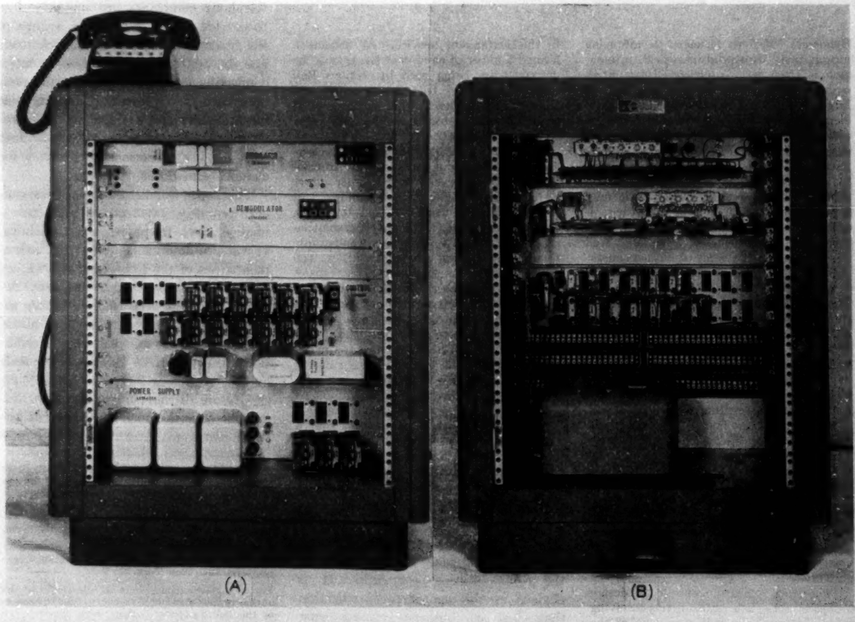


Fig. 1. Digital subset. A—Front view. B—Rear view

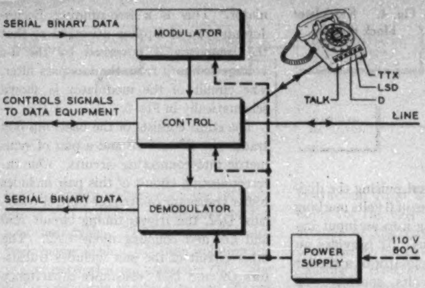


Fig. 2. Digital subset, block diagram

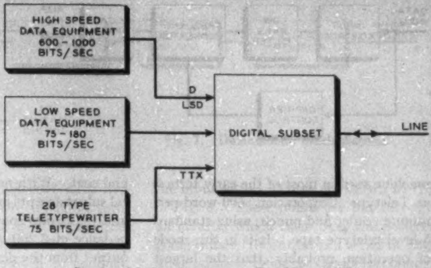

HIGH SPEED DATA EQUIPMENT 600 - 1000 BITS/SEC
LOW SPEED DATA EQUIPMENT 75 - 180 BITS/SEC
28 TYPE TELETYPEWRITER 75 BITS/SEC
D
LSD
TTX
DIGITAL SUBSET
LINE

Fig. 3. Modes of operation: D, LSD, and TTX

when a binary 0 or "space" is presented another voice frequency F_2 is transmitted. Selection of suitable frequencies F_1 and F_2 must take into consideration bandwidth restrictions on the channel, bit rates expected, and possible interference with signaling circuits connected to toll channels. An appropriate set of frequencies for use where the most restrictive narrow-band voice channels may be encountered is

$F_1 = 900$ cps

$F_2 = 1{,}400$ cps

If the subset is used with private line circuits where the bandwidth is less restrictive (say, 400 to 2,500 cps), then an appropriate set of frequencies is

$F_1 = 1{,}100$ cps

$F_2 = 1{,}900$ cps

Using the higher frequencies will, of course, allow higher bit rates. The subset allows individual adjustment of these frequencies for optimum operation over the facility used.

It is proposed to use the digital subset in the Dataphone service at a maximum speed of 600 bps. For data transmission over private-line channels the speed will be 750 bps. A higher rate of 1,000 bps may be obtained by special delay equalization of the private-line channel.

Fig. 1 shows front and rear views of initial models of the digital subset. The electronic circuitry is entirely transistorized. Fig. 2 shows a block diagram of the digital subset from which the various equipment elements in the photographs can be identified. The modulator takes the binary data from the data-producing machine and modulates it to frequencies F_1 and F_2 for transmission to the line; the demodulator receives signals from the line, limits and shapes them, and passes a detected signal to the associated data-reproducing machine in the form of serialized binary d-c pulses; the control circuit is a switching unit to allow association of a number of different data-producing reproducing equipments with the modulator and demodulator; the telephone set allows alternate use of the voice facility for speech purposes and data, and also controls the selection of the connected data equipment; the power supply services the modest power requirements of the transistor circuitry and the relays in the control panel.

Digital Subset Modes of Operation

Before discussing the operation of the transmission part of the digital subset, it will be convenient to present a general idea of the control function design. Three basic modes of operation are provided for; these are selected by operation of push buttons on the telephone set associated with the digital subset.

A teletypewriter mode (TTX) allows transmission from and to a 100-wpm (words per minute) *28*-type teletypewriter. In this mode appropriate connections are made from the modulator and demodulator to the *28* teletypewriter, and a low-pass filter suited to this low speed of transmission is switched into the demodulator. The TTX mode, of course, does not exploit the full speed capability of the voice channel. Its provision, however, is useful in a number of respects. Many business routines are now built around standard teletypewriter procedures. Thus, the digital subset may be used in basic business programs without essential changes in these programs. In addition, the unattended features (to be discussed in the following) of the digital subset rely on the remote-control signals available from a *28* teletypewriter. These features will be discussed. Inclusion of this mode of operation in the first digital subsets will also lead to an answer to the question, "Is this mode of operation useful to the business community?"

A low-speed data mode of operation (LSD) is also provided in the digital subset. This mode allows transmission at speeds up to about 180 bps. It is included to allow operation with automatic tape transmission machines such as the Commercial Controls Corporation Teledata machine. This mode is selected by operation of the LO SP DATA button on the telephone set. Operation of this button connects the data machine to the digital subset and selects the appropriate filter in the demodulator.

A high-speed data *D* mode of operation is also provided. Here the maximum capability of the digital subset is used. This mode is obtained by operation of the DATA button on the telephone set. As far as the characteristic distortion of the modulator-demodulator is concerned, bit rates considerably higher than 1,000 bps are possible in this mode of operation. However, the delay characteristics of speech channels generally limit the bit rate to about 1,000 bps. The higher rates are accompanied by increasingly large distortions so that the maximum capability is obtainable only by the use of exceedingly stable auxiliary synchronizing units. This type of synchronizing unit integrates over several transitions and provides a clock signal, so that the received signal is sampled precisely in the center of a bit. In this way transmission is satisfactory even where transmission distortion is rather severe.

Where narrow speech channels may be encountered (for example, "emergency bank" circuits) bit rates of 600 bps are realizable with distortion low enough so that simple synchronization such as start-stop may be employed.

It is expected that a number of different manufacturers' equipment will be usable with the digital subset in the higher speed data mode. A representative data

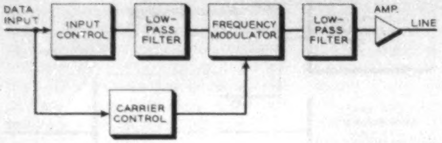


Fig. 4. Modulator block diagram

machine used in most of the early tests is a Teletype Corporation 600-word-per-minute reader and punch, using standard 5-level teletype tape. It is in this mode of operation, probably, that the largest economic advantage of Dataphone operation will be evident.

Fig. 3 shows in block form the three basic modes of operation of the digital subset. The input and output requirements in each of the modes are described in the next section.

Input and Output Signals

In the TTX and LSD modes of operation the digital subset expects to work with telegraph-type equipment. Therefore the subset is arranged to accept the usual contact closures and opens for mark and space signals. Information from the digital subset to the data machines is in the form of loop closures and opens; spacing, 0 milliamperes, marking, 20 milliamperes into a *255A* relay or equivalent.

For the D mode of operation either polar or neutral signals can be received and sent. With neutral pulsing the digital subset accepts pulses of 0 volts marking and −12 volts spacing into an input impedance of 2,200 ohms. It provides an output from the demodulator of a similar nature; marking 0 volts, spacing −12 volts into an impedance of 2,200 ohms. With polar pulsing the subset accepts pulses of +6 volts marking and −6 volts spacing into the same input impedance. The demodulator output is +6 volts marking and −6 volts spacing into an impedance of 800 ohms.

Digital Subset Modulator

A block diagram of the modulator is shown in Fig. 4. The input binary signal from the data-producing equipment is fed through an input control block where the various inputs are standardized. Following this, a low-pass shaping filter is introduced to shape the mark-space transitions to obtain proper keying of the frequency modulator. The output of the shaping filter is a voltage of one value for marking and another for spacing. This is fed to a 6-transistor multivibrator frequency modulator. This is a free-running Eccles-Jordan type of flip-flop arranged so that the frequency is governed by the d-c voltage fed to it from the low-pass filter. The circuit of the modulator is shown schematically in Fig. 5.

The stage consists of the basic flip-flop transistors *Q7* and *Q8* and a pair of symmetric interconnecting circuits. One interconnecting circuit of this pair includes buffer transistor *Q6*, trigger control transistor *Q12*, the trigger-timing circuit *R30* and *C6*, and coupling diode *CR2*. The other circuit of the pair includes transistors *Q9* and *Q11*, resistance-capacitance timing circuit *R29* and *C5*, and coupling diode *CR1*. A half-cycle of the operation of the circuit will be described.

Assume that transistors *Q7* and *Q12* are on. Diodes *CR1* and *CR2* are reverse-biased. Timing capacitor *C5* has, during the previous half-cycle been charged to the ON collector potential of *Q11*. This capacitor will discharge through *R29* from the ON potential of *Q11* toward the less positive keying voltage delivered from the input control circuit. Capacitor *C5* discharges until it reaches a voltage which forward-biases diode *CR1*. The more negative the keying voltage, the sooner this will occur, and the higher the frequency of the circuit. When *CR1* conducts, it conducts current away from the base of *Q7* and turns it off. When *Q7* turns off, the collector base cross coupling to *Q8* turns that transistor on. The collector drop in *Q8* is transmitted through emitter follower *Q9* to the base of *Q11*, turning it on. Transistor *Q11* provides a low-resistance charging path for *C5* which

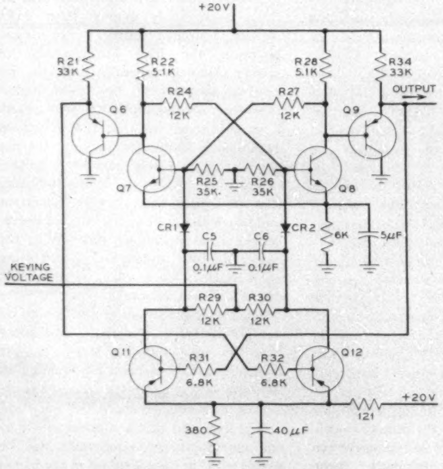


Fig. 5. FM modulator circuit

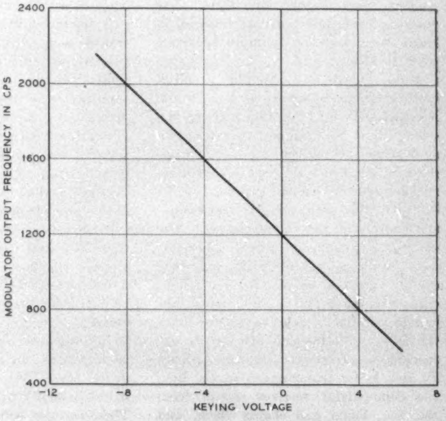


Fig. 6. Output frequency versus keying voltage

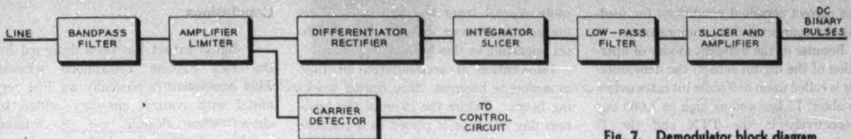


Fig. 7. Demodulator block diagram

is immediately recharged to the ON voltage of *Q11* collector preparatory to the next half-cycle. This voltage rise also back biases diode *CR1*, thus isolating the trigger network from transistor *Q7*. The latter previously has been turned off and the resultant rise in collector voltage has been coupled through *Q6* to turn off *Q12*. This results in unclamping *C6*, which will start to discharge toward the keying voltage for the next half-cycle. The circuit is symmetrical, so that the same action occurs during the second part of the cycle. A very linear frequency versus voltage characteristic is obtained from this stage; see Fig. 6.

The output of the modulator stage is sent through a harmonic suppression filter to eliminate all harmonics from the square wave generated, so that the line signal is essentially sinusoidal. A line amplifier is provided to couple this signal to the telephone line. The amplifier is capable of delivering a maximum output to the line of about +10 db above one milliwatt into 600 ohms. The normal output level will be about 0 db above one milliwatt at this point.

Notice that the block diagram of the modulator shows a block called carrier control; see Fig. 4. This is necessary to permit transmission of carrier only at appropriate times. Since the digital subset is connected to the telephone plant over a 2-wire telephone line, and since the same frequencies are used in both directions the carrier must only be turned on at the transmitting station. The carrier control senses mark-space transitions from the data equipment and turns the carrier on only when data is being transmitted. It is arranged so that when transmission is via teletypewriter keyboard in the TTX mode of operation, carrier is immediately transmitted when an attendant depresses a character on the keyboard. Carrier will remain on for about 300 milliseconds after the last mark-space transition so that during normal-speed typing, carrier is on continuously. On the other hand, if back-and-forth operation is desired, the other end of the circuit can seize control of the channel after this time interval has expired. The carrier control is also arranged to turn on carrier in response to an instruction from a connected data machine for the case where no rapid turnaround of the circuit is contemplated. The carrier control turns off carrier by causing a transistor switch to conduct, connecting collector and emitter of *Q8*; see Fig. 5.

Digital Subset Demodulator

A block diagram of the demodulator is shown in Fig. 7. The line signal is fed through a band-pass filter to an amplifier limiter. The limiter is designed to provide a standardized square-wave output of constant amplitude over about a 40-db range of input levels. It makes use of six transistors in two feedback amplifier combinations. The circuit is arranged so that no feedback is effective at lowest input levels. Diodes in the feedback path open at high levels, resulting in large feedback ratios. Additional shunt diode limiting is used at the highest levels. Fig. 8 shows typical waveforms at various places in the demodulator. The FM input to the digital subset is shown in (A) and the limiter output is shown on (B). Following the limiter is a differentiating circuit whose output consists of positive and negative pulses at the zero crossings of the input wave. These are shown at (C) in Fig. 8. These pulses are put through an inverter which turns over the positive going pulses; its output is a series of negative pulses at twice the carrier frequency; Fig. 8(D). These pulses trigger a transistor blocking oscillator which puts out negative pulses of about 15-microsecond width, each of which corresponds to a zero crossing of the original line signal. When the input is marking (lower frequency) these pulses are spaced farther apart than when the input is spacing (higher frequency). Following the blocking oscillator is an integrating stage. This stage charges a capacitor between pulses. If pulses are occurring at a high rate (spacing line signal) the capacitor in this stage does not have time to charge to its full potential before being reset by the next pulse. If the pulses occur at a lower rate this charging process will proceed to a higher voltage before being terminated. The next stage is a slicer which discriminates between these two charge levels. The waveform after slicing is shown in Fig. 8(E). The slicer output is averaged in a low-pass filter and it passes to a second slicer which produces sharp square waves at the desired bit rate as shown in (F). An output stage is used to provide

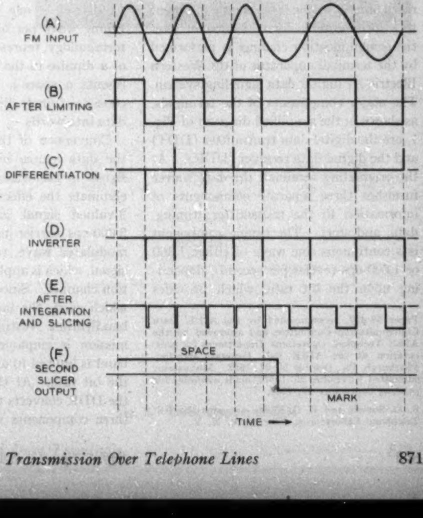


Fig. 8. Digital subset waveforms

the correct terminal conditions for feeding the connected data equipment.

Because of the different modes of operation of the digital subset, the demodulator is called upon to handle bit rates as low as about 75 bps and as high as 1,000 bps respectively in the TTX and the D modes of operation. Optimum low-pass filters differ for these different bit rates so that different filters are used under control of the mode selection relays in the control circuit.

A carrier detector block is shown in Fig. 7 in the demodulator. For control purposes this circuit gives an indication whenever carrier (either marking or spacing) is received by the demodulator. This might also be used to give an end of message indication to an attendant.

Unattended Operation

An attractive possibility in Dataphone service is remote delivery and pickup of data at a station which is unattended. The digital subset is arranged to allow operation of this sort in conjunction with *28ASR* teletypewriters equipped with remote control stunt box features. Tests are now under way with the digital subset operating in this fashion.

Information is accumulated on tape at a remote business office during working hours. Before the close of the business day the tape is placed in a suitable tape-transmitting device. During the evening hours a central record location reaches the unattended station by telephone in the usual way. The digital subset is arranged to answer the call automatically and to set itself into the TTX mode of operation. In this mode of operation the *28ASR* at the remote station is able to respond to special codes typed at the controlling point. For example, the attended station may type a code which tells the remote station to transmit the previously prepared message via some high-speed data machine and the digital subset. Receipt of this code will cause the digital subset to switch to the proper mode for this transmission and to start the tape machine sending. After the transmission the controlling point can initiate another action if required or can type a special code to cause disconnect.

Conclusions

A digital subset has been designed for the Bell System Dataphone service. This equipment is basically an FM terminal with control circuitry added to allow rather flexible use. A limited number of these terminals have been manufactured for trials in various fields of data transmission. It seems certain that the demands on the communication industry for better and more flexible means for transmission of data will multiply several times in coming years. The digital subset described in this paper is merely a beginning in this field of service.

References

1. BINARY DATA TRANSMISSION TECHNIQUES FOR LINEAR SYSTEMS, **M. L. Doelz, E. T. Heald, D. L. Martin.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., May 1957.

2. INFORMATION THEORY IMPACT ON MODERN COMMUNICATIONS, **Pierre Mertz.** *AIEE Transactions*, vol. 76, pt. I, Sept. 1957, pp. 431–37.

3. TRANSMISSION ASPECTS OF DATA TRANSMISSION SERVICE USING PRIVATE LINE VOICE TELEPHONE CHANNELS, **P. Mertz, D. Mitchell.** *Bell System Technical Journal*, New York, N. Y., Nov. 1957.

SAGE Data Terminals

R. O. SOFFEL
NONMEMBER AIEE

E. G. SPACK
NONMEMBER AIEE

IN THE SAGE (semiautomatic ground environment) System,[1,2] the function of converting the data in the form generated or used by the processing equipment to a form suitable for transmission over the communication channel is performed by the terminal apparatus of the Western Electric *A1* digital data signaling system. The major components of the terminals, as shown in the simplified diagram of Fig. 1, are the digital data transmitter (DDT) and the digital data receiver (DDR). At the originating terminal, the data source furnishes three separate components of information to the transmitter: timing, data, and start. The timing component is a continuous sine wave of either 1,300 or 1,600 cps (cycles per second), depending upon the bit rate, which provides means for synchronizing the system. Message information, contained in the data component, is in the form of dipulses. (A dipulse is one complete cycle of the timing wave per bit, which, in telegraph terminology, represents a mark. Absence of a dipulse in the data component represents a space.) The start component consists of single dipulses which group the data into words.

Conversion of the three outputs from the data source into a single, 3-valued signal is performed by the DDT. To eliminate the effect of carrier shift, this 3-valued signal is modulated with a 2,000-cps carrier to form an amplitude-modulated wave, referred to as the line signal, which is applied to the communication channel. Since some of the channels which carry the line signal have narrow bandwidths, vestigial side-band transmission is employed; the usable sideband is limited to approximately one half the bit rate. At the receiving terminal, the DDR converts the line signal into the three components which are faithful reproductions of the original information from the data source. These components are then fed into the data-using equipment.

Digital Data Transmitter

The DDT may be divided into four parts: timing, data, start and modulator circuits. Three of these parts act upon the information signals from the data source to form a single 3-valued signal. The fourth part, the modulator circuit, combines this signal with a carrier to produce the amplitude-modulated line signal, which is transmitted over the communication channel. These four sections are shown in block diagram form in Fig. 2. For the purpose of illustration, a segment of a word is shown which has its start pulse at t_0 and four marks or data bits in positions t_3, t_7, t_8, and t_9. In the SAGE System, the word length may vary up to 326 bits.

The three inputs to the DDT have the same amplitude, phase, and frequency. The timing component is applied through a narrow-band filter which reduces noise or any other distortion which may occur during transmission of the signal from the data source. After filtering, the signal passes through an adjustable phase shift circuit and is applied to a slicer where the

---

Paper 58-407, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted November 25, 1957; made available for printing September 17, 1958.

R. O. SOFFEL and E. G. SPACK are with the Bell Telephone Laboratories, New York, N. Y.

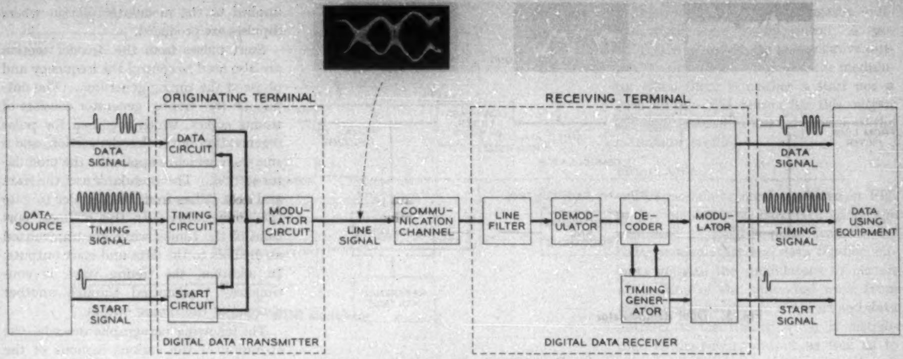

ORIGINATING TERMINAL
DATA SOURCE
DATA SIGNAL
TIMING SIGNAL
START SIGNAL
DATA CIRCUIT
TIMING CIRCUIT
START CIRCUIT
MODULATOR CIRCUIT
DIGITAL DATA TRANSMITTER
LINE SIGNAL
COMMUNICATION CHANNEL
RECEIVING TERMINAL
LINE FILTER
DEMODULATOR
DECODER
TIMING GENERATOR
MODULATOR
DATA SIGNAL
TIMING SIGNAL
START SIGNAL
DATA USING EQUIPMENT
DIGITAL DATA RECEIVER

Fig. 1. Simplified AI digital data signaling system

sine wave is converted into rectangular pulses. These pulses, in turn, are differentiated to form positive and negative spikes which are applied to a normally cut-off amplifier so that only the positive spikes cause it to conduct. The output of this amplifier consists of negative spikes, which are employed in the sampling gates of the data and start circuits. The adjustable phase shifter is used to center the sampling spikes.

In the data section, the data component of the input is sliced to convert the dipulses into rectangular pulse form. Two outputs are derived from the slicer, one when a data pulse is present, and the other for the no-data condition. These outputs are applied to the sampling gate and they steer the timing spikes to the inputs of the data flip-flop. The flip-flop always remains in the last state into which it was driven. Since the circuit is driven both on and off, the output must necessarily be an integral number of periods of the timing wave. The flip-flop responds to the timing spikes to produce negative-going rectangular pulses when data is present in the input signal.

Operation of the start circuit is similar to that of the data circuit. Start dipulses are sliced and the output is applied to a gate where it is combined with timing spikes to drive the start flip-flop. However, the positive output pulse of the flip-flop is used to indicate the presence of a start dipulse.

A summing network combines the outputs of the data and start circuits to produce a 3-valued signal. This composite signal is applied to the modulator through a low-pass filter which removes the frequency components above 1,000 cycles. In the modulator the composite signal is combined with a 2,000-cps carrier to form the line signal. The modulator output passes through a negative feedback amplifier and is then pre-equalized to compensate for delay and amplitude distor-

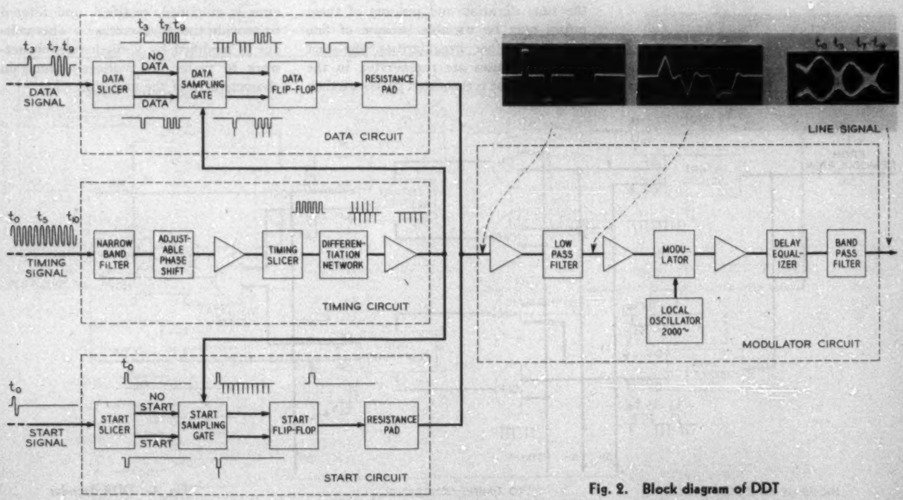

t_3 t_7 t_9
DATA SIGNAL
DATA SLICER
NO DATA
DATA
DATA SAMPLING GATE
DATA FLIP-FLOP
RESISTANCE PAD
DATA CIRCUIT
t_0 t_5 t_{10}
TIMING SIGNAL
NARROW BAND FILTER
ADJUSTABLE PHASE SHIFT
TIMING SLICER
DIFFERENTIATION NETWORK
TIMING CIRCUIT
t_0
START SIGNAL
START SLICER
NO START
START
START SAMPLING GATE
START FLIP-FLOP
RESISTANCE PAD
START CIRCUIT
t_0 t_3 t_7 t_9
LINE SIGNAL
LOW PASS FILTER
MODULATOR
LOCAL OSCILLATOR 2000~
DELAY EQUALIZER
BAND PASS FILTER
MODULATOR CIRCUIT

Fig. 2. Block diagram of DDT

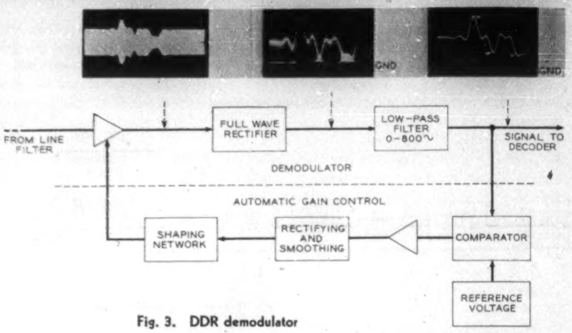

GND
GND
FROM LINE FILTER
FULL WAVE RECTIFIER
LOW-PASS FILTER 0-800∿
SIGNAL TO DECODER
DEMODULATOR
AUTOMATIC GAIN CONTROL
SHAPING NETWORK
RECTIFYING AND SMOOTHING
COMPARATOR
REFERENCE VOLTAGE

Fig. 3. DDR demodulator

tion in the entire transmission system. Extraneous modulation products are removed by filtering before the signal is applied to the communication channel.

Digital Data Receiver

As mentioned previously, the function of the digital data receiver is to reproduce the timing, data, and start components of the transmitted information from the line signal. To perform this function, the DDR must:

1. Detect the presence or absence of data and start pulses in the line signal.
2. Generate a timing wave which is synchronous with the original timing and which is properly phased, relative to the data signal.
3. Produce as outputs, data, and start dipulses which are faithful reproductions of the original signals from the data source and also a timing wave with the proper phasing.

The major sections of the DDR are shown in block diagram form as part of Fig. 1. Briefly, the receiver operates as follows: At the receiving terminal, the line signal is applied to a band-pass filter which provides the vestigial side-band shaping and also removes the out-of-band noise accumulated during transmission. The filtered signal is demodulated to recover the modulation envelope which consists of a wave containing positive and negative peaks representing the start and data pulses, respectively. These pulses are separated in proper sequence by the decoder section. Since the time duration and phasing of these pulses may be variable because of line noise and other irregularities, the start and data pulses are regenerated in the decoder. The regenerated pulses are then applied to the modulator section where dipulses are produced.

Start pulses from the decoder section are also used to control the frequency and phase of the timing generator. The output of the timing generator consists of timing spikes, which are used for pulse regeneration in the decoder section, and a sine wave which is applied to the modulator section. The modulator uses the start and data pulses from the decoder to gate the timing wave. In this manner, portions of the timing wave are transmitted as dipulses to the data and start outputs. In addition, the timing wave is continuously transmitted through another channel to the timing output.

The following paragraphs describe the operation of the various sections of the DDR in greater detail.

Demodulator

Fig. 3 shows the demodulator section in block diagram form along with the waveforms at various points in the circuit. In the demodulator, the filtered line signal is amplified, full-wave rectified, and filtered to recover the original 3-level signal. Positive peaks of the recovered envelope represent start pulses, while the negative peaks represent the data pulses. Since the proper operation of the remainder of the DDR circuit depends upon the signal being held to a close tolerance, an AGC (automatic gain control) system is employed. The peak of the start pulse is compared with a direct-voltage reference of approximately 45 volts. The difference is amplified, rectified, and filtered to provide the AGC voltage. This voltage is modified by a diode-resistor network to match the voltage-versus-gain characteristic of the amplifier.

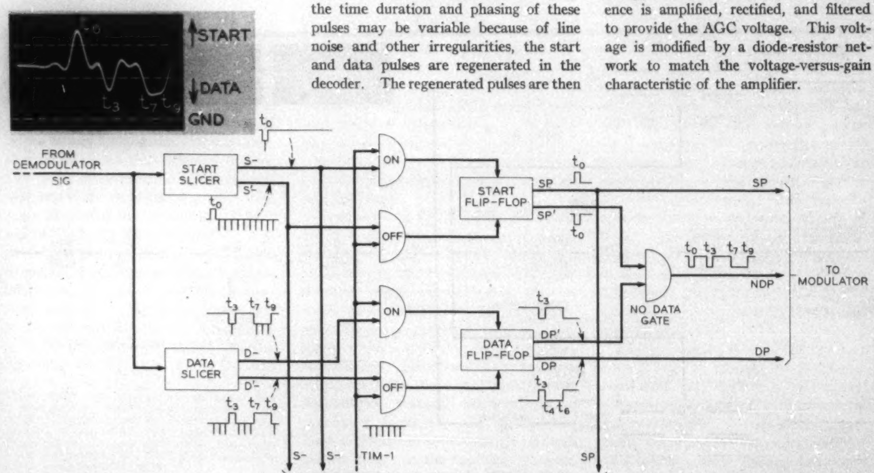

t0
START
DATA
t3 t7 t9
GND
t0
FROM DEMODULATOR
SIG
START SLICER
S-
S'-
t0
ON
OFF
t0
START FLIP-FLOP
SP
SP'
t0
SP
t0 t3 t7 t9
TO MODULATOR
NDP
NO DATA GATE
t3 t7 t9
ON
t3
DATA FLIP-FLOP
DP'
DP
DP
DATA SLICER
D-
D'-
t3 t7 t9
OFF
t3
t4 t6
S-
S-
TIM-1
SP
TO TIMING GENERATOR

Fig. 4. DDR decoder

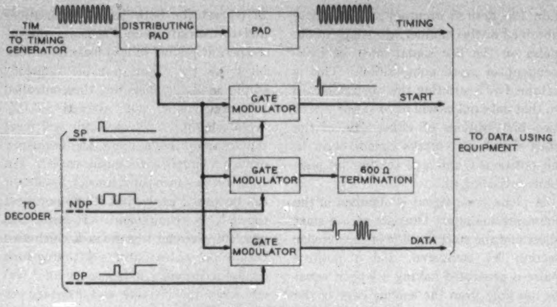

TIM
TO TIMING GENERATOR
DISTRIBUTING PAD
PAD
TIMING
GATE MODULATOR
START
SP
GATE MODULATOR
600 Ω TERMINATION
TO DATA USING EQUIPMENT
TO DECODER
NDP
GATE MODULATOR
DATA
DP

Fig. 5. DDR modulator

The peak of the start pulse, rather than the average carrier, is used to control the amplifier gain because the value of the average carrier is a function of word composition and thus may be continually changing. The value of the start peak, however, is fixed, and therefore it becomes a good reference point. Nevertheless, a considerable burden is placed on the AGC circuit since there may be more than 300 data bits between start pulses.

DECODER

Conversion of the 3-level output from the demodulator into data and start components takes place in the decoder, shown in Fig. 4. The modulator output is connected to a pair of slicers, one for recovering start pulses and the other for recovering the data pulses. These slicers are 2-stage differential amplifiers having an accurate, fixed bias corresponding to the mid-point of the desired slice. Use of two gain stages insures that the second stage will overload rapidly, resulting in a maximum slice thickness of about one volt. The slicer outputs are rectangular pulses. These pulses, however, are too greatly distorted to control the dipulse modulator directly. Therefore, in a manner similar to that described for the transmitter, the slicer outputs are used to control gates, which steer accurate timing spikes from the timing generator to control flip-flop circuits.

Both the start and data flip-flops are arranged to produce a positive output to indicate the presence of a pulse in the 3-level signal or a negative output when there is no pulse. A no-data gate combines the two negative outputs to provide an output to the modulator when there is neither a start nor a data pulse. The reason for this output will be explained in the description of the modulator section.

MODULATOR

The modulator section, shown in Fig. 5, is composed of balanced, gate-type modulators. Positive pulses from the decoder section start and data flip-flop circuits actuate the modulators to permit application of the sinusodial wave from the timing generator to the start and data outputs. This timing wave is distributed to the timing output as well as to the modulators. All three outputs must be held to the same level, within a fraction of a decibel. Therefore, a pad is added in the timing output to compensate for the loss in the start and data modulators. Since the input impedance of the modulator is very high when there is no gate pulse and is 600 ohms when a gate pulse is present, the termination presented to the distributing pad by these two modulators will vary with the composition of the 3-level signal. Such a variation could result in a modulation of the timing output unless, of course, the timing amplifier supplying the signal has a very low output impedance. The alternative to a low-impedance amplifier is to keep the termination of the distribution pad from varying. This has been accomplished

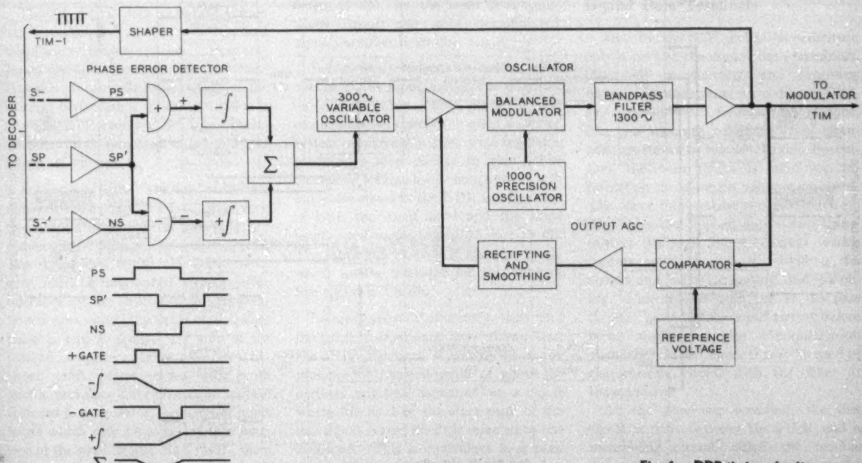

TIM-1
SHAPER
PHASE ERROR DETECTOR
TO DECODER
S-
PS
+
+
−∫
SP
SP'
Σ
S-'
NS
−
−
+∫
300 ∿ VARIABLE OSCILLATOR
OSCILLATOR
BALANCED MODULATOR
BANDPASS FILTER 1300 ∿
TO MODULATOR
TIM
1000 ∿ PRECISION OSCILLATOR
OUTPUT AGC
RECTIFYING AND SMOOTHING
COMPARATOR
REFERENCE VOLTAGE
PS
SP'
NS
+GATE
−∫
−GATE
+∫
Σ

Fig. 6. DDR timing circuit

by providing a dummy modulator which is controlled by the output of the no-data gate. The dummy modulator provides a 600-ohm termination when there are neither start nor data signals. As a result of this method of operation, the distribution pad is properly terminated at all times and the timing output is independent of word composition.

TIMING GENERATOR

The timing generator section may be divided into two main circuits a phase error detector and a variable oscillator. Block diagrams of these circuits are given in Fig. 6.

To provide a satisfactory error rate in the transmission of data over a noisy communication channel, the sampling spikes used in the regeneration of data must have a frequency exactly equal to the bit rate of the data source. In addition, the spikes must be accurately phased with respect to the line signal. The latter problem will be dealt with first. Because of the narrow transmission band provided by some of the communication channels, timing recovered from the data transitions cannot be employed to establish the phase, since word composition could cause the phase to vary by as much as ±90 degrees. The method used to establish the proper phase is based upon monitoring the sliced start pulse and insuring that the sampling spike is located at the mid-point of the pulse. To obtain the desired accuracy of the phase obtained in this manner, the shape of the pulse on the line signal must be independent of word composition. This is attained by restricting the word make-up so that data pulses will never occur within two bit positions of either side of the start signal (i.e., marks cannot occur in bit positions t_1 and t_2 or the two bit positions preceding t_0).

A phase error signal is obtained in the following manner. Outputs of the start slicer and the start flip-flop of the decoder section are compared, and a positive pulse is generated having a length equal to the time from the leading edge of the sliced start pulse to the leading edge of the start flip-flop pulses. The latter point corresponds to the occurrence of the sampling spike. Similarly, a negative pulse may be derived which has a duration equal to the time from the occurrence of the sampling spike to the trailing edge of the sliced start pulse. These two pulses are integrated to produce direct voltages of opposite sign which are proportional to the pulse durations. By adding the voltages in a summing network, the resulting algebraic sum provides an output signal which supplies both the magnitude and direction of the phase error. This signal is then used to control the frequency and phase of the oscillator section.

Requirements for the variable oscillator are quite severe. It must be easily varied to correct for both its own variations and the variations in the source timing. To permit pulling in and locking on 326-bit words, the variation in the frequency of the oscillator, including the controlled variation, must not exceed ±0.1%. While a tuning fork oscillator will meet the accuracy requirement, the frequency of such a circuit is not easily varied. An L/C (inductance-capacitance) oscillator can be varied easily, but it will not meet the ±0.1% requirement. To solve this problem, a circuit was devised which uses both types of oscillator. A tuning fork oscillator provides a frequency of 1,000 cps when the bit rate is 1,300 bits per second or 1,300 when the bit rate is 1,600 bits per second. A variable L/C oscillator provides a 300-cps frequency. The two frequencies are combined in a double balanced modulator and the desired component is recovered by filtering. As a result of this method of operation, the accuracy of the 300-cps oscillator must be held to ±0.4%.

The L/C tank of the variable oscillator is tuned to a nominal frequency of 300 cps, and the amplifier section of this oscillator is arranged so that its phase shift of 300 cps may be varied as a function of the phase error voltage. Since the net phase shift in the loop must be 360 degrees, the frequency of oscillation changes until this requirement is met.

Amplitude stability of the timing wave is obtained by comparing the timing wave

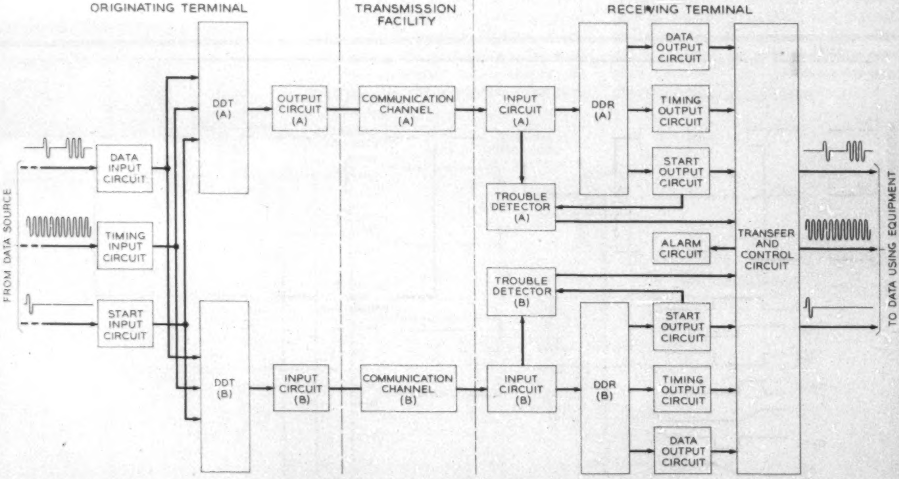

ORIGINATING TERMINAL
TRANSMISSION FACILITY
RECEIVING TERMINAL
FROM DATA SOURCE
DATA INPUT CIRCUIT
TIMING INPUT CIRCUIT
START INPUT CIRCUIT
DDT (A)
DDT (B)
OUTPUT CIRCUIT (A)
INPUT CIRCUIT (B)
COMMUNICATION CHANNEL (A)
COMMUNICATION CHANNEL (B)
INPUT CIRCUIT (A)
INPUT CIRCUIT (B)
TROUBLE DETECTOR (A)
TROUBLE DETECTOR (B)
DDR (A)
DDR (B)
DATA OUTPUT CIRCUIT
TIMING OUTPUT CIRCUIT
START OUTPUT CIRCUIT
ALARM CIRCUIT
START OUTPUT CIRCUIT
TIMING OUTPUT CIRCUIT
DATA OUTPUT CIRCUIT
TRANSFER AND CONTROL CIRCUIT
TO DATA USING EQUIPMENT

Fig. 7. Dual facility arrangement

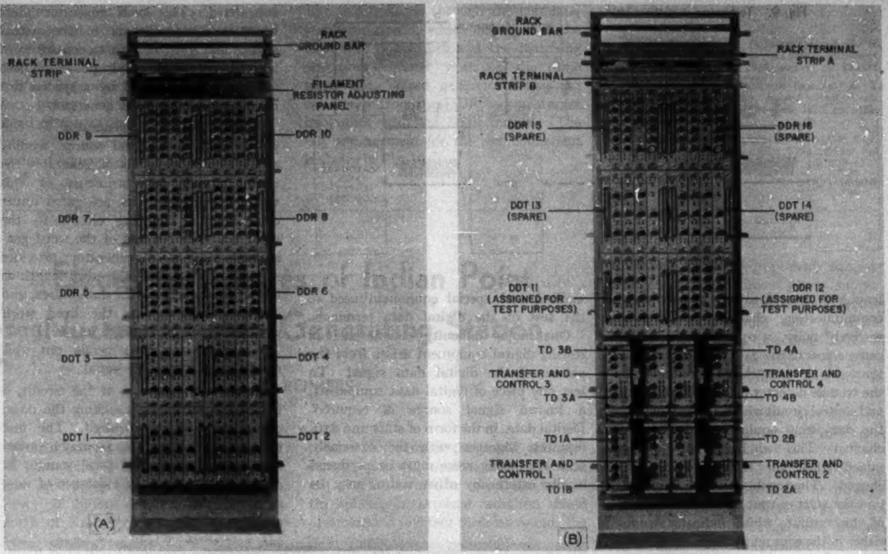

RACK TERMINAL STRIP
RACK GROUND BAR
FILAMENT RESISTOR ADJUSTING PANEL
DDR 9
DDR 10
DDR 7
DDR 8
DDR 5
DDR 6
DDT 3
DDT 4
DDT 1
DDT 2
(A)
RACK GROUND BAR
RACK TERMINAL STRIP A
RACK TERMINAL STRIP B
DDR 15 (SPARE)
DDR 16 (SPARE)
DDT 13 (SPARE)
DDT 14 (SPARE)
DDT 11 (ASSIGNED FOR TEST PURPOSES)
DDR 12 (ASSIGNED FOR TEST PURPOSES)
TD 3B
TD 4A
TRANSFER AND CONTROL 3
TRANSFER AND CONTROL 4
TD 3A
TD 4B
TD 1A
TD 2B
TRANSFER AND CONTROL 1
TRANSFER AND CONTROL 2
TD 1B
TD 2A
(B)

Fig. 8. A—DDT and DDR bay. B—Trouble detector and spare DDT and DDR bay

with a voltage reference provided by a zener diode. The different voltage is used to control the amplitude of the 300-cps component of the modulator.

GENERAL PERFORMANCE OF THE DDR

The *A1* digital data signaling system has been designed to work over standard telephone communication channels. James discusses these channels in detail.[3] However, to provide some details of the receiver performance, a brief discussion is given here of several of the transmission considerations connected with the use of telephone circuits.

Input signal level: The level of the line signal at the input to the DDR may vary as a result of differences in the net loss in connecting facilities, seasonal changes, etc. The signal level within the DDR, however, must be maintained within a close tolerance by the AGC circuit. As mentioned previously, the peak of the start pulse is used to control the gain of the receiver amplifier. Since the time between start pulses varies with word length, and since the gain control voltage is stored in a capacitor, the range of input levels which may be tolerated is a function of the word length. As a result, there exists a "word length-input product" which is somewhat analogous to the gain-bandwidth product of an amplifier. When the word length is 64 bits, for example, a range of input levels of 35 db (decibels) will cause a variation of only ±0.1 db in the demodulated signal. For a word length of 326 bits, the input level range which causes the same demodulated signal variation is 20 db.

Sudden level changes: A sudden level change in the input may occur when an automatic switch takes place in the communication channel or when a carrier system employing a pilot wire regulator produces a step change in gain. The amount of sudden level change which will not cause errors in the DDR is a function of both the input level and the word length, and varies from 0.75 to 1.75 db. The maximum theoretical sudden change which can be tolerated for this type of line signal is 1.8 db.

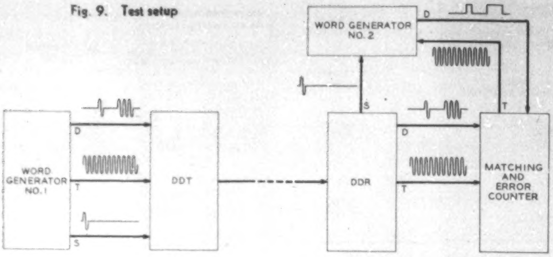
Induced noise: Laboratory tests and limited field experience have shown that the DDR functions satisfactorily in the presence of thermal noise of about 53 decibels adjusted measured at a point where the level of the start peak of the line signal is zero decibels referred to one milliwatt. This is equivalent to a peak signal to rms noise ratio of 37 db. Impulse noise peaks 18 db below the start peaks do not cause the DDR to exceed its allowable error rate. These values are reasonably close to the theoretical maximum allowable values.

Digital Data Terminals

Actually, the DDT and DDR constitute only a part of the digital data terminals. Complete transmitting and receiving terminal arrangements for a dual facility, i.e., a circuit which, for reliability reasons, has two separate communication channels, are shown in Fig. 7. In this illustration, the data source is supplying information to a single using equipment. The three information components from the source are applied to the two transmitters through input circuits which contain splitting pads for dividing the signals and jacks for testing and patching. Line signals from the DDTs pass through jacks in the output circuit before being applied to the communication channels. Either channel may be used as the working circuit with the other as the stand-by.

At the receiving terminal, the line signal is split between the DDR and a monitoring circuit called the trouble detector. One of the functions of the

Fig. 9. Test setup


WORD GENERATOR NO. 2
D
S
T
WORD GENERATOR NO. 1
D
T
S
DDT
DDR
D
T
MATCHING AND ERROR COUNTER

trouble detector is to determine when the communication channel becomes excessively noisy, a condition which could cause errors in the data. If, for example, the working channel becomes too noisy, the trouble detector actuates the transfer and control circuit which, in turn, switches the data-using equipment to the other channel. This switch will still be maintained when the trouble condition is cleared. The trouble detector also monitors the start output of the DDR. Loss of this output, which indicates trouble either in the communication channel or in the receiver, also results in the operation of the transfer and control circuit. When operating under control of the trouble detector, the transfer and control circuit will not switch to a circuit that is already in trouble.

The transfer and control circuit may also be operated manually at the using equipment. Associated with the circuit are lamps to indicate the status of the receivers (good or not good for service) and an audible alarm which signals the occurrence of a trouble condition.

Equipment arrangement for a digital data terminal is shown in Fig. 8(A) and (B). Since information is sent in both directions, this terminal has both transmitters and receivers. Fig. 8(A) shows a bay with four DDTs and six DDRs. The bay in (B) contains the DDT and DDR assigned for test purposes, spare transmitters and receivers, and four transfer and control circuits with the associated trouble detectors.

System Testing

The measure of effectiveness of any communications system is the ability of that system to transfer intelligence while preserving its form within prescribed limits. To maintain this effectiveness, some method of testing is required. The method used and the means by which it is performed is peculiar to the equipment involved. Fig. 9 shows the testing method and special equipment used in the test of the digital data terminals.

One major difficulty encountered in testing digital equipment arises from the nature of the digital data signal. To test any piece of digital data equipment, a known signal source is required. Digital data, in the form of start and data dipulses, together with the extremely accurate timing wave must be produced locally before any other testing may be considered. The local generator must simulate all conditions of the actual signals and must be able to control word length, amplitude and phase. These functions are provided by a special test circuit called the word generator.

When the word generator is employed as a test signal source, feeding a DDT, a local oscillator generates the timing wave, and segments of this wave, forming dipulses, are gated under control of a counting circuit to the start and data outputs of the word generator. The word generator provides means for controlling the word length as well as the location of the data pulses, and it continually repeats the fixed word pattern which is applied to the DDT. The transmitter then sends out this fixed pattern as the line signal.

At the receiving end of the circuit, a means for accurately checking the data, pulse by pulse, is required. The test equipment must be able to match pulses from two synchronized local sources of data and to record the existance of mis-

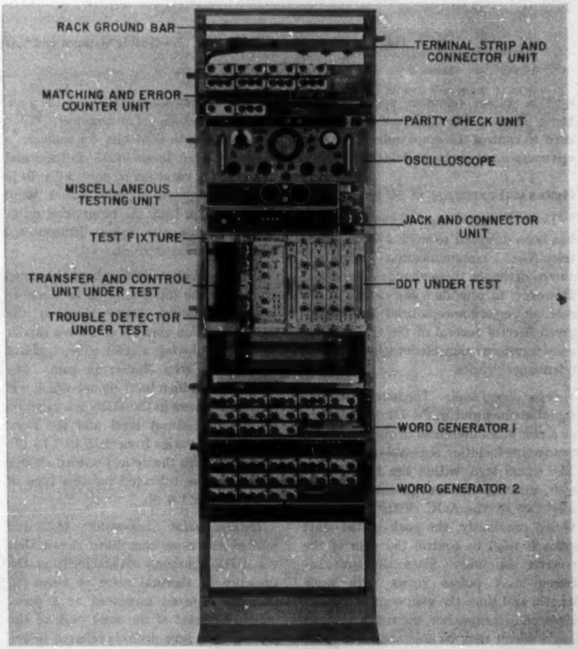

RACK GROUND BAR
TERMINAL STRIP AND CONNECTOR UNIT
MATCHING AND ERROR COUNTER UNIT
PARITY CHECK UNIT
OSCILLOSCOPE
MISCELLANEOUS TESTING UNIT
JACK AND CONNECTOR UNIT
TEST FIXTURE
TRANSFER AND CONTROL UNIT UNDER TEST
DDT UNDER TEST
TROUBLE DETECTOR UNDER TEST
WORD GENERATOR I
WORD GENERATOR 2

Fig. 10. Test frame

matches. These functions are performed by a second word generator (designated as word generator 2 in Fig. 9), and a matching and error counter.

Word generator 2 is set to have the same data pattern as word generator 1. Instead of generating its own timing, however, word generator 2 depends on the timing frequency generated by the DDR. The timing is fed through the matching and error circuit, which insures that the two word generators and the matching and error counter work on the same time base. Word generator 2 uses the start pulse from the DDR to synchronize its output with the DDR output. The bay arrangement for the test equipment is shown in Fig. 10.

References

1. SAGE Data System Considerations, R. G. Enticknap, E. F. Schuster. *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 824–32.

2. SAGE Data Transmission Service, A. E. Ruppel. *Bell Laboratories Record*, New York, N. Y., vol. XXXV, no. 10, Oct. 1957, pp. 401–05.

3. Communication Channels for SAGE Data Systems, R. T. James. *AIEE Transactions*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 838–43.

Electrical Features of Indian Point Nuclear Electric Generating Station

T. D. REIMERS
FELLOW AIEE

THE Consolidated Edison Company of New York is building a nuclear electric generating station on the Hudson River at Indian Point about 24 miles north of New York City. Fig. 1 is an aerial perspective sketch indicating the general arrangement of the completed plant.

The reactor is of the pressurized-water type using a mixture of highly enriched uranium oxide and thorium oxide as a fuel. It will operate at 1,500 psig (pounds per square inch gage) and at an average coolant temperature of 501 F (degrees Fahrenheit). The primary coolant system consists of four parallel loops which circulate demineralized light water between the reactor vessel and the steam generators. Two canned motor pumps circulate the water in each loop. The steam generators produce steam at 449 F and 405 psig. This steam will pass an oil-fired superheater to raise the temperature to 1,000 F and deliver 2,200,000 pounds of steam per hour to the turbine at this temperature and 355 psig.

The turbogenerator is a tandem-compound 1,800-rpm condensing unit with an estimated gross operating capacity of 275,000 kw.

The reactor, steam generators, and primary circulating pumps are housed in a pressuretight 160-foot-diameter steel sphere. To protect the public and personnel who will operate this unit and adjacent units in case of an accident in the sphere, an external radiation shield of concrete 3 to 6 feet thick surrounds the steel sphere.

Fig. 2 is an elevation showing the location of the main plant components.

Electrical Features Differing from Conventional Plant

It has been the aim of Consolidated Edison to make the design of the electrical features of this plant conform as closely as possible to that of their conventional plants. In most respects these features have been maintained.

However, particularly for the nuclear portion of the plant, there are certain requirements at the present state of reactor development which necessitate modification of conventional electrical designs. This paper deals particularly with these differences and the design modifications which result.

Reactor safety requirement suggests changes in the auxiliary power supply, control sources, controls, and instrumentation. To ensure safe operation, considerable reactor instrumentation is duplicated and triplicated. Very little instrument duplication is provided in the conventional plant.

Radioactivity, humidity, and temperature in the sphere present an insulation problem.

The large capacity of the generator with respect to a high-impedance tie through an adjacent company's system presents a stability and voltage-regulation problem under certain operating conditions.

Auxiliary Power Supply

For conventional unitized generating units Consolidated Edison has been providing one start-up source from an adjacent substation for the unit auxiliary power bus and has transferred this bus to the generator as a supply source as soon as this unit was synchronized to the system. In the event of the unit being automatically cleared from the system by operation of protective relays or boiler-turbine-generator sequence trips, the supply to the auxiliary bus becomes deenergized. Under this condition some lighting and a few small essential auxiliaries are automatically transferred to the start-up source.

In the reactor plant it is desirable to

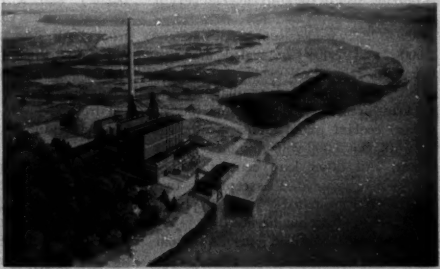


Fig. 1. Aerial perspective of Indian Point nuclear station

Paper 58-1194, recommended by the AIEE Nucleonics Committee and the Power Generation Committee of the Power Apparatus and Systems Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 30, 1958; made available for printing August 5, 1958.

T. D. Reimers is with the Consolidated Edison Company of New York, Inc., New York, N. Y.

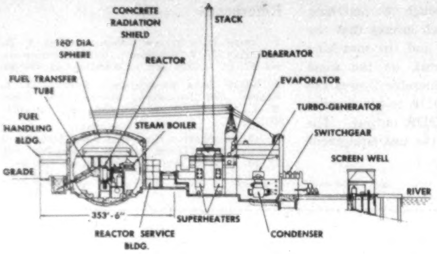

CONCRETE RADIATION SHIELD
STACK
160' DIA. SPHERE
REACTOR
DEAERATOR
FUEL TRANSFER TUBE
EVAPORATOR
FUEL HANDLING BLDG.
STEAM BOILER
TURBO-GENERATOR
SWITCHGEAR
SCREEN WELL
GRADE
RIVER
353'-6"
SUPERHEATERS
REACTOR SERVICE BLDG.
CONDENSER

Fig. 2. Elevation of Indian Point station

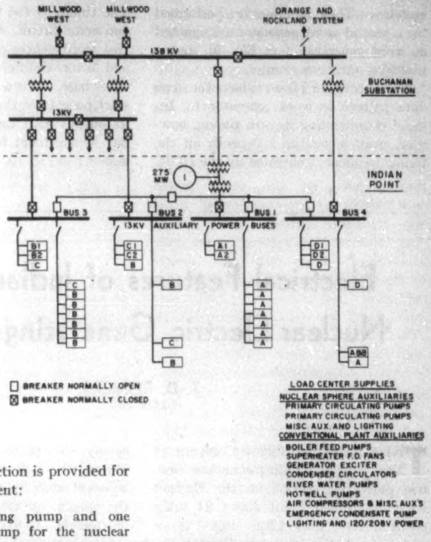

MILLWOOD WEST
MILLWOOD WEST
ORANGE AND ROCKLAND SYSTEM
138 KV
BUCHANAN SUBSTATION
13KV
275 MW
I
INDIAN POINT
BUS 3
BUS 2
BUS 1
BUS 4
13KV AUXILIARY POWER BUSES
B1
B2
C
C1
C2
B
A1
A2
D1
D2
D
C
B
B
B
B
B
B
C
B
A
A
A
A
A
A
A
A
D
ABB
A
BREAKER NORMALLY OPEN
BREAKER NORMALLY CLOSED
LOAD CENTER SUPPLIES
NUCLEAR SPHERE AUXILIARIES
PRIMARY CIRCULATING PUMPS
PRIMARY CIRCULATING PUMPS
MISC. AUX. AND LIGHTING
CONVENTIONAL PLANT AUXILIARIES
BOILER FEED PUMPS
SUPERHEATER F.D. FANS
GENERATOR EXCITER
CONDENSER CIRCULATORS
RIVER WATER PUMPS
HOTWELL PUMPS
AIR COMPRESSORS & MISC. AUX'S
EMERGENCY CONDENSATE PUMP
LIGHTING AND 120/208V POWER

Fig. 3 (right). Main connections and auxiliary power supply

maintain, at all times, supply to as much as possible of the reactor's electrically driven auxiliaries. To prevent possible reactor core damage in a "scram" operation, it is necessary to maintain service to at least one primary circulating pump, one seal make-up pump, one boiler-feed pump, electrical control and instrumentation sources, other small essential equipment, and essential lighting.

To provide this reliability the 13-kv auxiliary power bus has been designed in four sections. One bus is supplied from the generator terminals, two busses are each supplied by a feeder from different bus sections of an adjacent distribution substation, and the fourth bus is supplied through a step-down transformer from a 138-kv tie feeder to the neighboring Orange and Rockland Utilities, Inc., system.

Fig. 3 is a one-line diagram of the main connections and the auxiliary system.

Normally, the four 13-kv auxiliary bus sections are operated with the bus tie circuit breakers open. When the normal supply to a bus section becomes de-energized, that section may be tied manually to another section. No automatic transfer is provided for the 13-kv busses.

Each 13-kv light and power bus section has one feeder supplying load in the nuclear portion of the plant and one feeder supplying load in the conventional part of the plant.

These 13-kv feeders supply 937-, 1,250-, 2,000-, and 2,500-kva sealed dry-type transformers in load center substations located close to the large auxiliary motors or load areas. Each 1,600-hp (horsepower) coolant pump motor is supplied from its own 2,300-volt grounded-Y load center. All other load centers supply 440 volts grounded Y except for two 125/216-volt grounded-Y units supplying lighting and small auxiliaries.

Automatic throwover from its normal load center source to a source fed from a different 13-kv bus section is provided for the following equipment:

One fresh-water cooling pump and one river-water cooling pump for the nuclear plant cooling requirements.
Generator-exciter motor-generator set.
Valve motors.
Essential turbine shutdown auxiliaries.
One battery charger of each d-c control voltage system.
Main turbine auxiliary oil pump.
Fire pumps.

Remote manual transfer is provided for the following:

One spent-fuel cooling pump.
One process transfer pump.
One decay-heat cooling pump.
One seal and make-up pump.
One pressurizer heater group.
One containment cooling fan.
One control rod hydraulic pump motor.

Control and Instrumentation Supply

In Consolidated Edison's conventional plants control power is supplied at 125 volts d-c from two busses each having a battery and a rectifier or a motor-generator set. Control centers are fed from both busses through interlocked automatic throwover switches.

The 125-volt 60-cycle circuits for instrumentation are supplied from a bus fed from two 125/216-volt load centers through an interlocked automatic throwover switch.

In the nuclear plant much of the reactor instrumentation requires an exceptionally reliable 125-volt a-c supply for electronic circuits used in monitoring reactor operation. Severe voltage dips or an interruption of even a cycle to this a-c instrumentation supply would cause "scramming" of the reactor.

To provide this reliable supply, 3-unit motor-generator sets have been specified each having a normal drive motor supplied from one of the 440-volt 3-phase 60-cycle load centers, a d-c generator connected to a 125-volt d-c battery bus, which acts as a motor when the normal a-c drive supply becomes de-energized, and a 125/216-volt 3-phase 60-cycle generator.

The connections of these sets are shown in Fig. 4. Each of two of the sets normally supplies separate 125-volt d-c control and 125/216-volt 60-cycle instrument busses. The third set which is supplied from an essential 440-volt bus can be connected to either of the d-c control and a-c instrument busses through remote-operated selector switches.

Normally, these busses will be operated as shown in the figure. The spare motor-generator set may be connected to either one of the d-c control busses *DC-1* or *DC-2* and the instrument supply busses *60-1* or *60-2*. The respective busses would be tied only in the extreme emergency when the spare set is the only one available.

Important 125-volt d-c control centers will have a supply from each of the two control busses connected to it through an interlocked automatic throwover. The momentary interruption occurring during this transfer does not affect the operation of any equipment connected to this system since no normally energized equipment is supplied from this system.

Half of the nuclear instrument circuits requiring a reliable supply will be fed from each of the two 125-volt 60-cycle instrument busses. In the event that the normal a-c motor driving the associated a-c generator is de-energized, the set will continue to be driven from the battery of the 125-volt d-c bus by means of the d-c generator now operating as a motor. In such instances, the battery can usually be relieved by connecting the spare motor-generator set to the affected busses. In the extreme condition of losing the 440-volt supply to all the motor-generator sets, the reactor will have been scrammed by the safety protective system, and only a few instruments will have to be kept in service to monitor the conditions in the reactor core. By manually switching off all other circuits, the essential monitoring equipment can be operated by one of the a-c generators driven from the battery for at least 24 hours.

The reactor rod control system requires a 400-cycle power supply for magnetic amplifiers. This supply must also be exceptionally reliable because its loss for only 5 milliseconds will cause the reactor to scram. This supply is obtained from a 3-unit motor-generator set with a flywheel consisting of two 440-volt 60-cycle motors, each fed from a different auxiliary power source, driving a 400-cycle 120-volt generator. The loss of either normal 60-cycle supply source can be tolerated without affecting the 400-cycle power supply. A normally shutdown spare set is installed for this system.

A 48-volt d-c system is used for alarm circuits and for remote operation of switchgear.

Nuclear Instrumentation

The nuclear instrumentation differs from conventional boiler instrumentation in several ways:

1. The primary variable is neutron density in the reactor, as compared with fuel-firing rate in a conventional boiler.
2. The range of the variable is from $10^{-9}\%$ to 150% of power level, as compared with 1% to 100% for a boiler furnace.
3. The nature of the variable displayed and recorded is log power level and linear power as compared with temperature, steam flow, etc.

Another important distinction is that the nuclear instrumentation provides the information required for the control and the safety system, while on conventional furnaces the control and the safety systems are entirely separate.

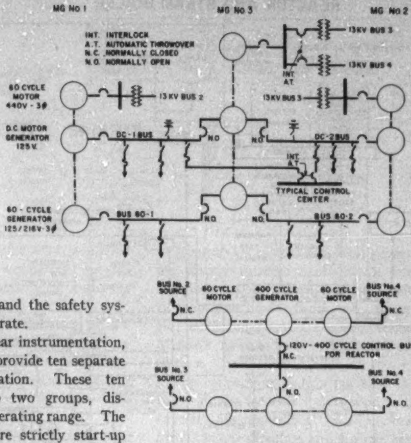

MG No 1
MG No 3
MG No 2
INT. INTERLOCK
A.T. AUTOMATIC THROWOVER
N.C. NORMALLY CLOSED
N.O. NORMALLY OPEN
13KV BUS 3
13KV BUS 4
INT.
A.T.
60 CYCLE MOTOR 440V - 3ϕ
13 KV BUS 2
13KV BUS 3
D.C MOTOR GENERATOR 125V.
DC-1 BUS
N.O
N.O
DC-2 BUS
INT.
A.T.
TYPICAL CONTROL CENTER
60 - CYCLE GENERATOR 125/216V-3ϕ
BUS 60-1
N.O.
N.O.
BUS 60-2
BUS No. 2 SOURCE
60 CYCLE MOTOR
400 CYCLE GENERATOR
60 CYCLE MOTOR
BUS No. 4 SOURCE
N.C.
N.C
120V- 400 CYCLE CONTROL BUS FOR REACTOR
N.C.
BUS No. 3 SOURCE
BUS No. 4 SOURCE
N.O.
N.O.
N.O.

Fig. 4. D-C control: 60-cycle instrumentation and 400-cycle control systems

The proposed nuclear instrumentation, shown in Fig. 5, will provide ten separate channels of information. These ten channels divide into two groups, distinguished by their operating range. The first four channels are strictly start-up ranges covering the power level from start-up neutron source ($10^{-9}\%$) to 0.1% of full load.

The other six channels overlap the upper end of the start-up range of power level and extend to 150% of full power. These are divided into two groups of which three channels provide log-level and period* information from $10^{-3}\%$ to 100%, while the other three channels display linear power level from 0.1% to 150%.

There is intentional duplication and 2-decade overlapping because of the extreme importance of this instrumentation to both the operators and the safety system. Thus, channel *2* duplicates channel *1*, both operating from neutron counters and providing indication of log count rate and period, which operate indicating instruments and recorders and, through an "auctioneering" device for log count rate and period, provide a signal from the higher of the two channels for optional use in the safety system. The auctioneer-

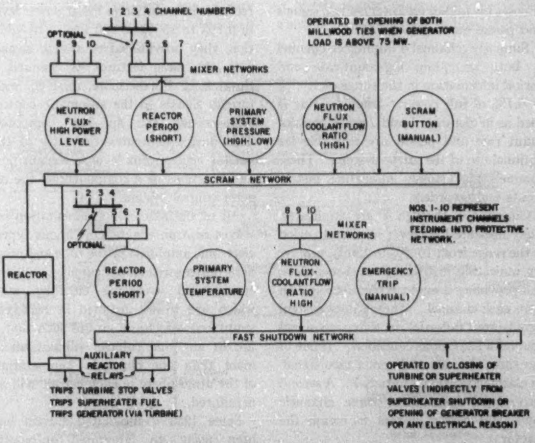

1 2 3 4 CHANNEL NUMBERS
OPTIONAL
8 9 10
5 6 7
MIXER NETWORKS
OPERATED BY OPENING OF BOTH MILLWOOD TIES WHEN GENERATOR LOAD IS ABOVE 75 MW.
NEUTRON FLUX-HIGH POWER LEVEL
REACTOR PERIOD (SHORT)
PRIMARY SYSTEM PRESSURE (HIGH-LOW)
NEUTRON FLUX COOLANT FLOW RATIO (HIGH)
SCRAM BUTTON (MANUAL)
SCRAM NETWORK
1 2 3 4
5 6 7
OPTIONAL
8 9 10
MIXER NETWORKS
NOS. 1-10 REPRESENT INSTRUMENT CHANNELS FEEDING INTO PROTECTIVE NETWORK.
REACTOR
REACTOR PERIOD (SHORT)
PRIMARY SYSTEM TEMPERATURE
NEUTRON FLUX-COOLANT FLOW RATIO HIGH
EMERGENCY TRIP (MANUAL)
FAST SHUTDOWN NETWORK
AUXILIARY REACTOR RELAYS
TRIPS TURBINE STOP VALVES
TRIPS SUPERHEATER FUEL
TRIPS GENERATOR (VIA TURBINE)
OPERATED BY CLOSING OF TURBINE OR SUPERHEATER VALVES (INDIRECTLY FROM SUPERHEATER SHUTDOWN OR OPENING OF GENERATOR BREAKER FOR ANY ELECTRICAL REASON)

Fig. 5. Reactor protection

---

* Period is a measure of the time in seconds for the power to change by the factor ϵ(2.718).

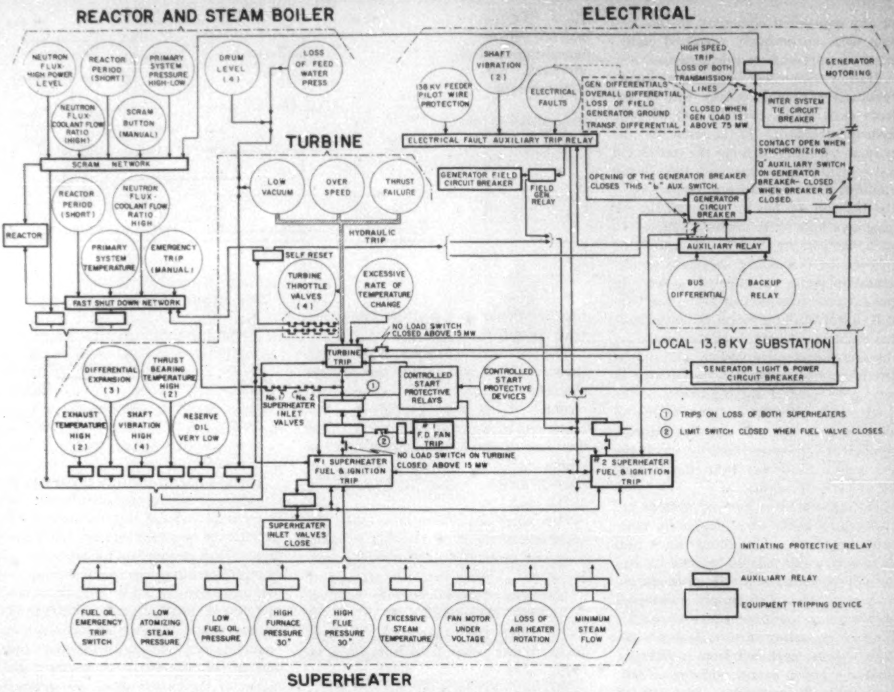

REACTOR AND STEAM BOILER
NEUTRON FLUX-HIGH POWER LEVEL
REACTOR PERIOD (SHORT)
PRIMARY SYSTEM PRESSURE HIGH-LOW
NEUTRON FLUX-COOLANT FLOW RATIO (HIGH)
SCRAM BUTTON (MANUAL)
DRUM LEVEL (4)
LOSS OF FEED WATER PRESS
SCRAM NETWORK
REACTOR
REACTOR PERIOD (SHORT)
NEUTRON FLUX-COOLANT FLOW RATIO HIGH
PRIMARY SYSTEM TEMPERATURE
EMERGENCY TRIP (MANUAL)
FAST SHUT DOWN NETWORK
TURBINE
LOW VACUUM
OVER SPEED
THRUST FAILURE
HYDRAULIC TRIP
SELF RESET
TURBINE THROTTLE VALVES (4)
EXCESSIVE RATE OF TEMPERATURE CHANGE
NO LOAD SWITCH CLOSED ABOVE 15 MW
TURBINE TRIP
DIFFERENTIAL EXPANSION (3)
THRUST BEARING TEMPERATURE HIGH (2)
EXHAUST TEMPERATURE HIGH (2)
SHAFT VIBRATION HIGH (4)
RESERVE OIL VERY LOW
No. 1 No. 2 SUPERHEATER INLET VALVES
CONTROLLED START PROTECTIVE RELAYS
CONTROLLED START PROTECTIVE DEVICES
①
②
#1 F.D. FAN TRIP
NO LOAD SWITCH ON TURBINE CLOSED ABOVE 15 MW
#1 SUPERHEATER FUEL & IGNITION TRIP
#2 SUPERHEATER FUEL & IGNITION TRIP
SUPERHEATER INLET VALVES CLOSE
ELECTRICAL
138 KV FEEDER PILOT WIRE PROTECTION
SHAFT VIBRATION (2)
ELECTRICAL FAULTS
GEN. DIFFERENTIALS
OVERALL DIFFERENTIAL
LOSS OF FIELD
GENERATOR GROUND
TRANSF. DIFFERENTIAL
HIGH SPEED TRIP LOSS OF BOTH TRANSMISSION LINES
CLOSED WHEN GEN LOAD IS ABOVE 75 MW
INTER SYSTEM TIE CIRCUIT BREAKER
GENERATOR MOTORING
ELECTRICAL FAULT AUXILIARY TRIP RELAY
GENERATOR FIELD CIRCUIT BREAKER
FIELD GEN RELAY
OPENING OF THE GENERATOR BREAKER CLOSES THIS "b" AUX. SWITCH
CONTACT OPEN WHEN SYNCHRONIZING
"a" AUXILIARY SWITCH ON GENERATOR BREAKER- CLOSED WHEN BREAKER IS CLOSED
GENERATOR CIRCUIT BREAKER
AUXILIARY RELAY
BUS DIFFERENTIAL
BACKUP RELAY
LOCAL 13.8 KV SUBSTATION
GENERATOR LIGHT & POWER CIRCUIT BREAKER
① TRIPS ON LOSS OF BOTH SUPERHEATERS
② LIMIT SWITCH CLOSED WHEN FUEL VALVE CLOSES
INITIATING PROTECTIVE RELAY
AUXILIARY RELAY
EQUIPMENT TRIPPING DEVICE
FUEL OIL EMERGENCY TRIP SWITCH
LOW ATOMIZING STEAM PRESSURE
LOW FUEL OIL PRESSURE
HIGH FURNACE PRESSURE 30"
HIGH FLUE PRESSURE 30"
EXCESSIVE STEAM TEMPERATURE
FAN MOTOR UNDER VOLTAGE
LOSS OF AIR HEATER ROTATION
MINIMUM STEAM FLOW
SUPERHEATER

Fig. 6. Protection system of nuclear station

ing device is an electronic device which chooses the higher (or lower) of two signals and passes on the chosen signal.

Similarly, channel *4* duplicates channel *3*, both providing log-count-rate and period information in the range of 10^{-8}% to 0.1% of full power. Auctioneering is used as in channels *1* and *2* and signals of count rate and period are available for optional use in the safety system. These channels also supply indicating instruments and recorders.

Channels *5* through *7* are triplicates. They measure log power level and period in the range from 10^{-3}% to 100%, supplying indicating instruments and recorders, and providing a signl to the safety system from each channel. These safety system signals are "fail-safe" in that zero signal denotes a hazardous condition. If one of the three channels delivers a zero signal, an alarm indication is initiated. A second zero signal from these three channels causes the safety system to scram the reactor.

Channels *8* through *10* are also triplicates. They provide indicated and recorder information of linear power level in 0.1% to 150% of full load. In addition, they provide safety system signals with the same features as channels *5* through *7*. Furthermore, these channels furnish signals to the automatic control system and the flux-flow computer. This device computes the ratio of the reactor neutron flux level to coolant flow and functions as a component of the reactor control system.

All of the reactor instrumentation involves neutron counters of various types, electronic amplifiers of the tube and magnetic designs with power supplies, auctioneering, and averaging circuitry, etc., which are to be installed in enclosed, ventilated cubicles in groups such that a fire in any one cubicle will not affect more than one channel. Temperature of the atmosphere in each cubicle will be monitored.

Since 1953 Consolidated Edison has been using an integrated automatic trip protection system for turbine generators and boilers, which was described 5 years ago in an AIEE paper.[1] This protection has proved very satisfactory in detecting undesirable conditions not immediately recognizable by the operator and in reducing the extent of damage. The system has proved very reliable and false tripping has been at a minimum.[2]

In the instance of the nuclear plant, essentially the same automatic protection will be provided for the turbine and superheater as would be provided normally for the turbine and boiler. Fig. 5 shows the additional protection for the nuclear reactor and its interconnection with the superheater, turbine, and generator protection.

Fig. 6 shows the over-all automatic protection of reactor, superheater, and turbine generator.

Insulation Problems in Reactor Sphere

Electrical insulating materials in the nuclear sphere will be subjected to conditions considerably more severe than in conventional power plants.

Ambient temperatures of 45 to 65 C (degrees centigrade) are expected to exist normally in the nuclear sphere, as compared with 40 to 50 C in the conventional plant. Humidity conditions in the sphere will be more severe and may at times approach 100% relative humidity.

Electric equipment in the sphere will be normally subjected to radiation of 50 to 150 roentgens per hour, a problem which does not exist at all in conventional plants.

All power, control, and instrumentation cables entering the sphere must be so installed as to prevent leakage at operating pressures up to 25 psig which may occur during an accident in the reactor.

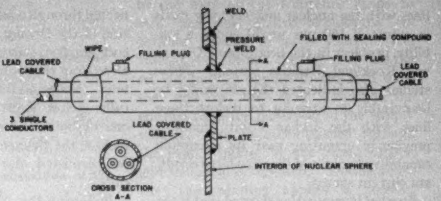
To minimize difficulties associated with electrical insulating materials under these more severe conditions, every effort has been made to place electric equipment outside the phere. This will also aid in maintenance since the radiation level prohibits employees from entering the sphere under operating conditions. All major contactors, switches, and similar equipment are located outside both the sphere and concrete radiation shield. Limit switches and terminal boards which have to be located in the sphere are of glass-base materials with silicone impregnants. Such devices generally will be in watertight housings. There are some 75 electric motors in the sphere ranging from fractional to 1,600 hp.

All motors in the sphere will be totally enclosed and will have class *H* insulation for maximum resistance to temperature and radiation effects. The primary circulating pump motors are of the canned type.

Motor circuits will be glass-tape silicone-rubber-impregnated lead-covered cable. The cable will be wiped to glands at the gasketed motor terminal box. Cables of 2,300 volts such as used for the 1,600-hp coolant pump motors and other large 440-volt cables will be terminated at the sphere wall in solder seal potheads whose metal body is welded to the sphere wall with the porcelain section outside the sphere and with the lead sheath of the cable wiped to the pothead gland inside the sphere.

Low-voltage small power and control cables will be lead covered and will be brought into the sphere through wiping glands as shown in Fig. 7. The several insulated conductors will be sealed in this sleeve with a suitable filling compound which hardens, fills in the spaces, and seals to both the individual conductor insulation and to the sleeve. Investigations and tests are now in progress to select the most suitable wire insulation material and filling compounds for this purpose and to withstand the radiation and temperature effects.

Fig. 7. Cable penetration through sphere


WELD
PRESSURE WELD
FILLED WITH SEALING COMPOUND
WIPE
FILLING PLUG
FILLING PLUG
LEAD COVERED CABLE
LEAD COVERED CABLE
A
A
3 SINGLE CONDUCTORS
LEAD COVERED CABLE
PLATE
INTERIOR OF NUCLEAR SPHERE
CROSS SECTION A-A

Both the potheads and the smaller wiping sleeves which penetrate the sphere must hold the 25-psig pressure which may be obtained in the sphere in a reactor accident. They will be designed for twice this pressure.

Outside the sphere the cable insulation for 2,300- and 440-volt power cables and for control cables will be ozone-resisting compound rubber as used in present stations.

Connection of Nuclear Generator to System

Fig. 8 shows the connections of the Indian Point nuclear electric generating plant to the Consolidated Edison and neighboring systems. The unit feeds at 138 kv into the new Buchanan substation located about three quarters of a mile from the generator. This substation has two 138-kv ties to Consolidated Edison's Millwood substation and one 138-kv tie connected to Orange and Rockland Utilities 69-kv system at Lovett generating station through a step-down autotransformer. Note that there is another tie loop from Lovett through the Orange and Rockland–Central Hudson–Niagara Mohawk systems back to Millwood substation.

The connection of this machine at a point on the systems having two large-capacity low-impedance ties on common towers in one direction and a single high-impedance tie through smaller systems in the other direction presents some operating problems should the Buchanan-Millwood lines become opened for any reason.

Either of the Buchanan-Millwood lines alone are adequate for the transmission of the 275-mw (megawatt) capacity of this new nuclear-powered generator to the system. If both of these lines are out of service, only about 75 mw of the total output can be absorbed locally and transmitted satisfactorily through the high-impedance ties of the Orange and Rockland–Central Hudson–Niagara Mohawk systems.

Network analyzer studies indicate that the simultaneous tripout of both Millwood

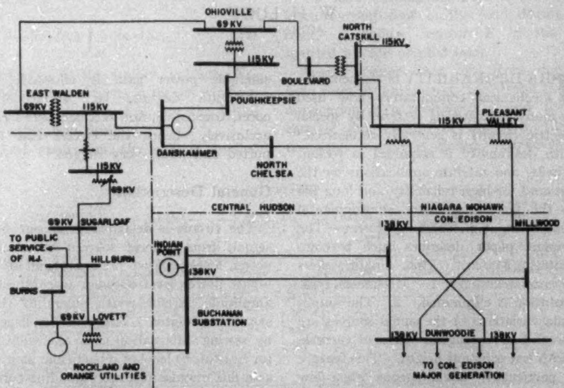

OHIOVILLE
69 KV
115 KV
NORTH CATSKILL
115KV
BOULEVARD
EAST WALDEN
69KV
115KV
POUGHKEEPSIE
115 KV
PLEASANT VALLEY
DANSKAMMER
NORTH CHELSEA
115KV
69KV
CENTRAL HUDSON
NIAGARA MOHAWK
CON. EDISON
69KV
SUGARLOAF
138KV
MILLWOOD
TO PUBLIC SERVICE OF N.J.
HILLBURN
INDIAN POINT
138KV
BURNS
69KV
LOVETT
BUCHANAN SUBSTATION
138KV
DUNWOODIE
138KV
ROCKLAND AND ORANGE UTILITIES
TO CON. EDISON MAJOR GENERATION

Fig. 8. Interconnection of Indian Point nuclear station to adjacent systems

lines with the nuclear unit carrying loads in excess of 150 mw would create a stability problem in the neighboring Orange and Rockland Utilities system. These studies also indicated that it would be inadvisable to attempt to reclose these lines with the Indian Point generator previously operating near its maximum capability because of the very large resultant current surges.

Further investigation showed that if the Indian Point generator was also tripped on the occurrence of a double Buchanan-Millwood tie outage, under peak load conditions and the outage of even a small Lovett generator, the neighboring system could not support both the Buchanan substation and the Indian Point auxiliary essential load.

The probability of obtaining the combination of the foregoing conditions is remote, but the requirement of a reliable supply to the essential Indian Point auxiliaries has led to the adoption of the following protective scheme:

Under any condition where the Indian Point generator output exceeds 75 mw, an opening of both of the Millwood ties will also scram the reactor and trip the generator and the Orange and Rockland tie breaker at Buchanan. This leaves the essential nuclear auxiliary load connected through a transformer on the feeder side of the Orange and Rockland tie circuit breaker, since this feeder remains alive from the Orange and Rockland system. Millwood tie or ties will automatically reclose if the fault has cleared. If one or both Millwood lines successfully reclose, the Orange and Rockland tie will be permitted also to reclose automatically.

Following this it becomes necessary to restart the reactor, bring the turbine generator up to speed, and resynchronize to the system.

In the event that the Millwood ties lock out on reclosure, the Orange and Rockland tie will not reclose automatically. The restoration of the Buchanan load, the ties and generator, will then be performed manually as load and capacity conditions permit.

Conclusions

1. The electrical features of a nuclear-powered turbine-generator plant are similar in many respects to a conventional plant.

2. To protect the reactor from core damage, a more diversified auxiliary supply is required than for conventional plants.

3. Very reliable sources, unaffected by system disturbances, of 125-volt 60-cycle power for instrumentation and of 125-volt 400-cycle power for the safety and rod control systems are required for the nuclear plant.

4. Over-all reactor-superheater-turbine-generator automatic protection uses the same principles developed for the conventional plant with added information obtained from the reactor.

5. Location of this large unit at an interconnection point having large low-impedance ties to one system and small high-impedance ties through another system presents some unusual protection problems.

References

1. PROTECTION OF TURBINE GENERATORS AND BOILERS BY AUTOMATIC TRIPPING, **H. A. Bauman, J. M. Driscoll, P. T. Onderdonk, R. L. Webb.** *AIEE Transactions*, vol. 72, pt. III, Dec. 1953, pp. 1248–60.

2. FIVE YEARS EXPERIENCE ON THE CONSOLIDATED EDISON SYSTEM WITH PROTECTION OF TURBINE GENERATORS AND BOILERS BY AUTOMATIC TRIPPING, **W. C. Beattie, H. A. Bauman, J. M. Driscoll, P. T. Onderdonk, R. L. Webb.** *Ibid.*, vol. 77, pt. III, 1958 (Feb. 1959 section) (*Paper No. 58–115*).

3. THE CONSOLIDATED EDISON INDIAN POINT NUCLEAR POWER PLANT, **G. R. Milne, F. R. Ward.** *Mechanical Engineering*, New York, N. Y., Mar. 1957 (condensed).

4. THE PROPOSED CONSOLIDATED EDISON COMPANY OF NEW YORK NUCLEAR POWER PLANT, **G. R. Milne, W. T. Moore.** World Power Conference, Vienna, Austria, June 17–23, 1956.

5. THE CONSOLIDATED EDISON NUCLEAR ELECTRIC GENERATING STATION, **O. G. Hanson, F. R. Ward.** Nuclear Science Congress, Chicago, Ill., Mar. 1958. (Available from American Institute of Chemical Engineers, New York, N. Y.)

Vol.
Pt. 1
195

A Magnetic-Amplifier Commutating and Pulse-Width Encoding Circuit

W. H. LUCKE
AFFILIATE AIEE

THE DESIRABILITY of replacing the mechanical commutators now used in many telemetering systems by special electric circuits is generally recognized.[1] This desirability is enhanced in rocket, missile, and satellite applications by the demand for high reliability and long life in the face of stringent environmental conditions, e.g., high g forces. The present paper describes such a commutating circuit. This circuit offers several advantages: 1. Mechanical commutation is eliminated. 2. The amplitude variations of the signal sources are encoded as rectangular pulses of variable width but constant height. This feature is particularly advantageous when low signal-to-noise ratios are encountered in the communication link. 3. Considerable power gain is afforded. 4. Solid-state devices, i.e., square-loop cores, transistors, and rectifiers, are used exclusively, hence the circuit can be potted and made very rugged.

General Description

The circuit is designed to accept d-c signals from several sources at a low power level and to present them at a higher power level encoded as constant-amplitude variable-width pulses to the system modulator. This is accomplished by storing information as to the voltage (or resistance) level of a particular source as a flux change in a corresponding core. Subsequently this information is "read-out" of the core as a rectangular pulse of variable width, which is applied directly to the modulator input.

The power source for the circuit is a transistor saturable-core multivibrator, hereafter referred to as the supply. During negative half-cycles of the supply, all signal sources are "sampled" simultaneously, i.e., the flux level in each core is caused to change by an amount proportional to the level of its signal source. On the following positive half-cycle, the output of the supply is applied to each core in turn, driving it to saturation. By Faraday's law the product of this constant voltage and the time it takes the core to saturate is equal to $\Delta\phi$, the flux change caused by the signal source. Prior to saturation the core acts as a transformer; hence the output of a core consists of the transformed supply output voltage existing for the time it takes the core to saturate. Since each

---

Paper 58-1162, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted April 24, 1958; made available for printing July 25, 1958.

W. H. LUCKE is with the U. S. Naval Research Laboratory, Washington, D. C.

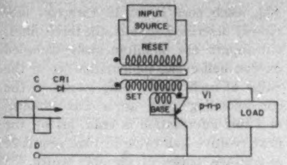


Fig. 1. The basic unit of the sequential set circuit

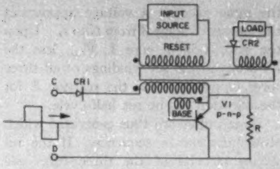


Fig. 2. An alternative configuration of the basic unit

core goes to saturation (i.e., sets), in turn it is clear that the output of the circuit consists of a sequence of variable-width pulses.

It is evident that the key to circuit operation is the means by which cores which are "reset" (i.e., have information stored in them) in parallel are set each in its turn. This is accomplished by a novel arrangement of simple transistorized magnetic-amplifier stages such that, upon the application of the positive half-cycle of the supply, each stage sets in turn, all other stages being quiescent. This basic arrangement (hereafter called the sequential set circuit) forms a flexible unit to which various means of set and reset can be applied, and may be used to form the basis of a variety of circuits for generating a train of pulses.

It should also be mentioned that the circuit is capable of giving a considerable power gain. The power level required of the signal sources is determined by the energy required to reset the core, whereas readout, or set, utilizes energy from a source furnished for that purpose and may be at a high power level.

The Sequential Set Circuit

In the following description of the operation of the sequential set circuit, reference to the sources and application of set and reset voltages will be made in general terms only. This approach has the two-fold object of keeping the description simple and focusing attention on the basic circuit and its possibilities. A description of the means of set and reset will be given later in connection with the discussion of the commutator as a whole.

The basic unit, or stage, of the sequential set circuit is diagrammed in Fig. 1. An alternating square-wave supply voltage is applied to the power input terminals *C* and *D*, which feed the set winding and the p-n-p transistor and load in parallel through the rectifier *CR1*. (The rectifier is necessary to block current in the set windings during reset.) The core is reset from the input source by an amount proportional to the voltage of the source. The source is presumed to present an open circuit to the reset windings during set. The dots are the conventional winding polarity markers.

Now suppose that the core has been reset by an amount $\Delta\phi$ from $+\phi_r$ to some arbitrary ϕ_c. When terminal *C* becomes positive with respect to terminal *D*, magnetizing current will flow through the set winding and the load. However, as soon as flux starts to change in the core, a voltage will be induced in the base drive winding in such a direction as to make the base of the p-n-p transistor negative. With proper design, this base voltage is large enough to hold the p-n-p transistor in saturation, i.e., emitter-to-collector resistance a minimum, for currents equal to the magnetizing plus base-load current which must be passed by the set winding. In the transistor saturated condition, the base-to-emitter voltage is very low—of the order of a few tenths of a volt. Hence, the voltage across the load is a few tenths of a volt; the majority of the current flows through the transistor. This situation holds until the core saturates at a time *t* determined by the relation

$$\int_0^t V_{CD}dt = N_{\text{set}}\int_{\phi_c}^{\phi_s} d\phi = \Delta\phi \qquad (1)$$

where V_{CD} is the supply voltage, N_{set} is the number of turns in the set winding, ϕ is flux, and ϕ_c and ϕ_s are the initial and saturation fluxes.

Upon saturation, the voltage applied to the base of the transistor drops to zero and the emitter-to-collector resistance rises to a large value. The voltage V_{CD} minus the *IR* drop in the set windings is applied to the load until the end of the cycle. During the ensuing negative half-cycle of the supply voltage, reset is accomplished through the reset winding by the signal source, and the cycle is repeated. During reset, the base winding applies a positive voltage to the base of the transistor. This has the effect of further increasing the emitter-to-collector resistance so that the transistor is effectively an open circuit. Current flow in the load is prevented by virtue of the blocking of the rectifier by V_{CD}, which is now negative and must, at all times, be equal to or greater than the voltage induced in the set winding by the resetting voltage, V_{reset}. A silicon-junction diode gives an added margin of safety because of its higher conduction threshold

Since

$$\Delta\phi = \frac{1}{N_{\text{reset}}}\int_0^{T/2} V_{\text{reset}}dt,$$

where N_{reset} is the number of turns of the reset winding and $T/2$ is the half-period of the supply voltage. The greater V_{reset}, the greater $\Delta\phi$. By equation 1 the greater $\Delta\phi$, the greater *t*, and the later in the set half-cycle saturation occurs. Hence output bears a complementary relation to input.

An alternative configuration, in which output bears a direct relation to input, is diagrammed in Fig. 2, where the reset and set circuits are unchanged, but the load has been connected across a winding provided for that purpose. The rectifier in the load winding has been added to block current flow in the load during reset. A ballast resistor *R* is substituted for the original load.

Again assume that the core has been reset by an amount $\Delta\phi$. Upon the application of a positive V_{CD} and the ensuing flux change in the core, the transistor switch is closed. However,

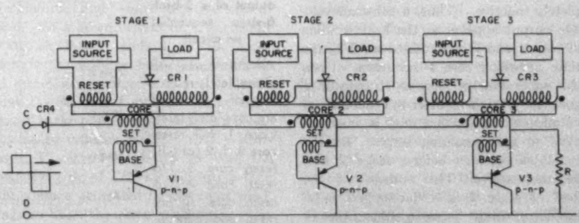


Fig. 3. Simplified version of a 3-stage sequential set circuit

here the base drive must be great enough to hold the transistor in saturation for magnetizing plus transformed load current. In this case, the transformed voltage V_{CD} is applied to the load for the time t for saturation to occur and is then removed. After saturation, V_{CD} is applied to the resistor R for the duration of the set half-cycle. The presence of R is necessary to prevent high-frequency oscillation after saturation. This oscillation is caused by the flux in the core dropping from its saturation to its remanance value. This sudden flux change in combination with the capacitance of the windings causes ringing which is amplified by the transistor to the point at which sustained oscillation occurs. By fixing R in conjunction with V_{CD} to give a current about equal to that flowing before saturation, the difficulty may be eliminated.

Evidently, circuit *2* has disadvantages not suffered by circuit *1:* 1. the transistor must pass the transformed load current, and, 2. power is wasted by R. In certain situations, however, these disadvantages may not be great.

The sequential set circuit proper consists of a series combination of the two basic circuits just described. A simplified 3-stage version, as illustrated in Fig. 3, will be described to introduce the basic mode of operation. Again it is assumed that reset has just been completed. Upon the application of the positive half-cycle of V_{CD}, there is a short transient burst of current and a small reversible change of flux in all cores;[2] the emitter-to-collector resistance of all transistors is momentarily lowered. However, the lowering of this resistance in the transistor of stage *1* means that less current and voltage is available for the succeeding stages. The condition is quickly established in which the first transistor is saturated, which means that only a few tenths of a volt of potential is available for causing flux change in the succeeding cores. This voltage is too small to cause magnetizing current to flow, so the last two stages are completely inactive. Thus, a constant-voltage output appears at the load winding of core *1* and is maintained until t_1, the time at which core *1* saturates. Upon the saturation of core *1*, transistor *1* goes to a high emitter-to-collector resistance value, so that V_{CD} is now applied to the remaining stages. By the same transient as before, stage *2* now becomes active. The voltage on the load of stage *2* is a flat-topped pulse extending from t_1 to t_2, the time of saturation of core *2*. By repetition of the same process, a voltage appears at the output of stage *3* from t_2 to t_3. Upon the saturation of core *3*, V_{CD}, less the IR drops in the set windings of all three cores, appears across the resistor R for the duration of the set half-cycle.

Three pulses are thus generated which always appear in sequence. If the resetting circuits of the three cores are independent of each other, and they may be made so by a means to be presently described, these three pulses may be made independent of each other in duration. The sole interdependence is that pulse *2* begins at the instant pulse *1* ends and pulse *3* starts upon the termination of pulse *2*.

When the supply voltage V_{CD} goes negative, the rectifier *CR4* blocks (this will be discussed more fully later on), as do the load rectifiers, and for the duration of the negative half-cycle the cores are reset by the input sources. On the subsequent positive half-cycle of V_{CD}, each core is set in turn, as previously described. Thus, the three different inputs are simultaneously sampled on one half-cycle, information as to the value of each input being stored in the corresponding core as a flux change, and this information is read out on the next positive half-cycle in time sequence as voltage pulses of variable width.

If the input sources represent voltages of continuously variable amplitude, the output pulses are continuously variable in width. The possibility of constructing a long delay circuit with a variable delay is worth noting. If an output were taken from the third core only, then this output could be shifted in time readily by varying the reset of cores *1* and *2*. In fact, the circuit could be simplified by combining cores *1* and *2* into a single core of large volt-second capacity.

The amplitudes of the output pulses are not flexible, since they are all derived

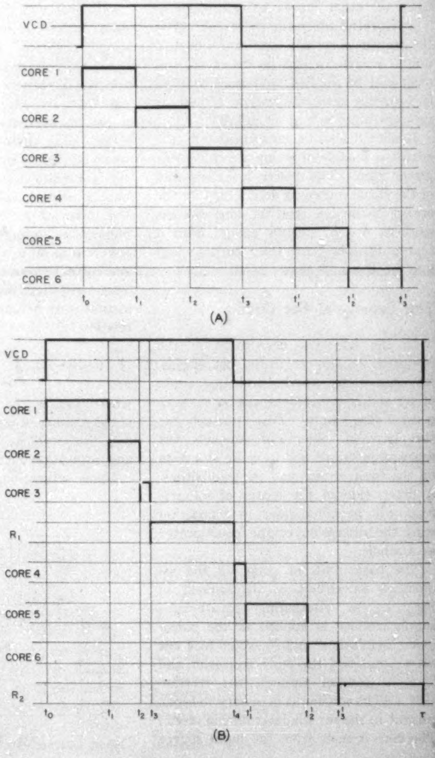


Fig. 4. Idealized output of a 2-bank 6-stage sequential set circuit

A—All cores are completely reset
B—Not all cores are completely reset. Core 1, full reset; core 2, 1/2 (of full) reset; core 3, 1/6 reset; core 4, 1/6 reset; core 5, full reset; and core 6, 1/2 reset

from V_{CD}. However, by varying the turns ratios of the load and set windings the pulses may be set at predetermined, but different, levels. In designing such a circuit, it is apparent that the total volt-seconds of the cores as viewed from the set windings must be less than, or equal to, the integrated volt seconds in one half-cycle of the supply voltage V_{CD}, i.e.,

$$\int_0^{T/2} V_{CD}dt = N_{S1}\Delta\phi_1 + N_{S2}\Delta\phi_2 + N_{S3}\Delta\phi_3 + \int_{t_3}^{T/2} I_sRdt,$$

where N_{S1} is the number of turns in the set windings of core *1*, N_{S2} the set turns on core *2*, etc., the $\Delta\phi$'s represent change in flux of the cores, and I_s is the current which V_{CD} causes to flow in the ballast resistor R. Or, replacing the fluxes by their volt-second equivalents,

$$\int_0^{T/2} V_{CD}dt = \int_0^{t_1} V_{CD}dt + \int_{t_1}^{t_2} V_{CD}dt \times \int_{t_2}^{t_3} V_{CD}dt + \int_{t_3}^{T/2} I_sRdt$$

By careful design, the last integral on the right may be made zero for the conditions of maximum reset of all three cores.

The objection that the sequential set circuit just described has a half-cycle of "dead" time may be eliminated by the addition of a second bank of three cores arranged to be set on the negative half-cycles of V_{CD}. The construction in so far as the set and base drive windings are concerned is identical with that of the first bank. The only difference is that the input from the supply is reversed. The first bank is set during positive half-cycles of the supply while simultaneously the second bank is being reset. On the negative half-cycles, the first bank is reset while the second bank is set. In Figs. 4(A) and 4(B), the outputs for such a double-bank arrangement are diagrammed for 1. the condition in which all cares are reset the maximum amount, giving a total volt-second value equal to that of the supply, and 2. in which core *1* has been completely reset, core *2* has been one-half reset, core *3* one-sixth reset, core *4* one-sixth reset, core *5* fully reset, and core *6* one-half reset.

Fig. 4(B) shows an inherent characteristic of a circuit in which the period of the supply is fixed and the reset of the cores is independently varied. After the application of the volt-seconds corresponding to a particular scheme of reset, the remainder of the volt-seconds in the supply half-cycle appears across the ballast resistor. Ordinarily, this interval represents useless "dead" time. However, situations might well arise in which use could be made of this pulse which occurs only after all stages have saturated.

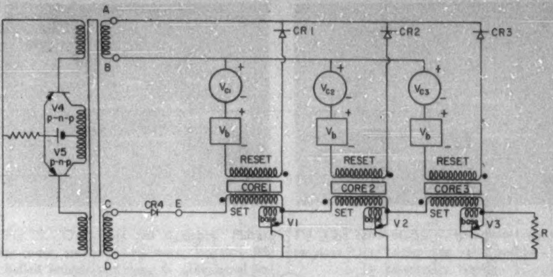


Fig. 5. Simplified 3-stage sequential set circuit showing method of accomplishing independent reset

The Commutating and Encoding Circuit

Various methods of accomplishing reset are possible. Fig. 5 shows a version of the method used in a 14-channel telemeter commutating-encoding system which will be described in detail later.

The supply source is a transistor saturable-core multivibrator. The circles labeled V_{c1} and V_{c2} are signal sources and may be either resistances or voltages. The rectangles labeled V_b are bias voltages (they may also be resistances) which were found necessary for this configuration. The need for them is made clear in the Appendix. The rectifiers *CR1*, *CR2*, and *CR3* will be discussed presently. The remainder of the circuit is identical with that of Fig. 3, except for the loads and load windings, which have been omitted for simplicity. The dots shown at the various winding ends are the conventional polarity indicators. Current into a dotted end takes the core toward positive saturation, current into an undotted end toward negative saturation. At a given instant on a particular core, all dotted ends have the same polarity. On the set half-cycle, terminals *A* and *C* are positive with respect to *B* and *D*. As core *1* sets, a voltage is induced in the reset winding of such a polarity as to unblock rectifier *CR1*. However, V_{AB} is set by design to hold the negative side of *CR1* more positive than the voltage generated by the reset winding, so that the reset winding is effectively isolated and can neither load the set circuit nor interact with any of the other resetting circuits. At the same time, rectifiers *CR2* and *CR3* keep V_{AB} from affecting cores *2* and *3*, since in these two stages there is no voltage induced in the reset windings. If the signal and bias sources are voltages, their polarity is in a direction opposite to the reset winding voltages; i.e., they help V_{AB} block the rectifiers. Since the voltage induced in any reset winding is zero under either 1. the condition of no current in the set winding, or 2. the condition of current in the set winding but with the core saturated, it is apparent that regardless of which cores are saturated, which is firing or which are yet to be fired, the reseting circuits are isolated from one another and from the voltage V_{AB}, except for rectifier leakage. With good semiconducting diodes, the leakage currents can be held to values well below magnetizing current, so that their effect is negligible.

When the set half-cycle is completed, the voltages V_{AB} and V_{CD} are reversed. V_{CD} is blocked by the rectifier *CR4*; V_{AB}, however, is in such a direction as to unblock rectifiers *CR1*, *CR2*, and *CR3*. This unblocking puts the three resetting circuits in parallel across V_{AB}. Resetting currents flow simultaneously from the bus at *B* through the control sources, the bias sources, the reset windings and the rectifier to the bus at *A*.

The fact that the cores are reset in parallel means that simultaneous voltages are induced in the set windings. Unfortunately, these voltages are in such a direction as to add, and if care is not exercised in design, their sum may be of sufficient magnitude to unblock rectifier *CR4* and throw a load on the resetting circuit. Suppose that the turns ratio of the set-to-reset windings is one, that $V_{AB} = V_{CD}$, and that all control and bias sources are zero, then the voltage at

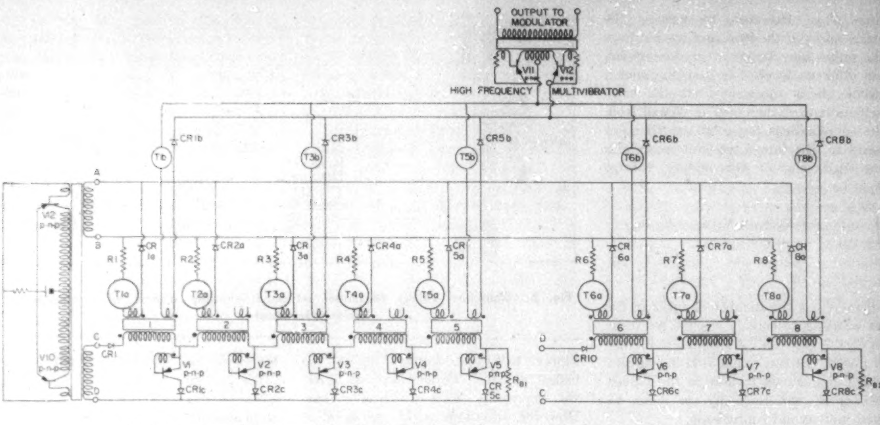

OUTPUT TO MODULATOR
V11
V12
HIGH FREQUENCY MULTIVIBRATOR
T1b
CR1b
T3b
CR3b
T5b
CR5b
T6b
CR6b
T8b
CR8b
A
B
C
D
V12 p-n-p
V10 p-n-p
R1
R2
R3
R4
R5
R6
R7
R8
CR1a
CR2a
CR3a
CR4a
CR5a
CR6a
CR7a
CR8a
T1a
T2a
T3a
T4a
T5a
T6a
T7a
T8a
1
2
3
4
5
6
7
8
CR1
CR10
V1 p-n-p
V2 p-n-p
V3 p-n-p
V4 p-n-p
V5 p-n-p
V6 p-n-p
V7 p-n-p
V8 p-n-p
CR1c
CR2c
CR3c
CR4c
CR5c
CR6c
CR7c
CR8c
RB1
RB2

Fig. 6. Fourteen-channel encoder and commutator based on a 2-bank 8-stage sequential set circuit

the point *E* would be the sum of the individual voltages induced in the set windings, or $-3V_{CD}$, while the voltage at *C* would be $-V_{CD}$; clearly the rectifier would unblock.

In fact, there are three conditions which must simultaneously be satisfied for proper operation of the circuit: 1. The sum of the voltages induced in the set windings during reset cannot be greater than the supply voltage V_{CD}; 2. The total volt-seconds reset of the cores which takes place during the entire reset half-cycle must be less than or equal to the total volt-second capability of the supply voltage during the set half-cycle. 3. The voltage induced in the reset windings during the set half-cycle must not be large enough to unblock the reset circuit rectifiers.

These three conditions are stated analytically in the Appendix, and from them are derived the necessary design equations.

Telemetering Circuit

The circuit diagrammed in Fig. 6 shows an application of the sequential set circuit to telemeter encoding. Here a 2-bank sequential set circuit has been combined with a relatively high-frequency magnetic core multivibrator to obtain a 14-channel system.

The supply multivibrator operates at a frequency of 75.5 cycles per second, delivering a V_{CD} of 2.5 volts and a V_{AB} of 12 volts. The resistors *R1*, *R2*, ..., *R8* are the padding resistors referred to in the Appendix. The transducers *T1a*, *T2a*, ..., *T8a* are the control sources governing reset. In the circuit tested, they were resistors. The rectifiers marked *CR1a*, *CR2a*, ..., *CR8a* are the diodes used to isolate the reset windings during set. The rectifiers, *CR1c*, *CR2c*, ..., *CR8c* are diodes added in the collector circuits to prevent stage interaction during reset. Suppose, for example, that core *4* is not to be reset at all, but that cores *3* and *5* are to be reset the full amount. The positive voltage appearing at the undotted end of core *5* can feed through the resistance of R_{B1} to the bus connected to point *D*. In the absence of the rectifier *CR4c*, this positive voltage would be in the forward direction of the diode formed by the collector and base of *V4*, and resultant current flowing through the fourth stage base drive winding could partially reset core *4*. The positive voltage at the undotted end of core *3* would aid this effect because of back leakage through *V3*.

The high-frequency multivibrator drives the modulator. Built to operate in the range of 2 to 5 kilocycles, it takes advantage of the fact that, within limits, the fundamental frequency of such a multivibrator depends linearly on the voltage applied. The power sources are the load windings of stages 1, 3, 5, 6, and 8 feeding through the transducers (in this case, resistance) *T1b*, *T3b*, *T5b*, *T6b*, and *T8b*. The voltage applied to the high-frequency multivibrator is given by

$$V_{\text{app}} = \frac{N_L}{N_s}\left(V_{CD} - V_{T(b)} - I_L r_{(b)}\right)$$

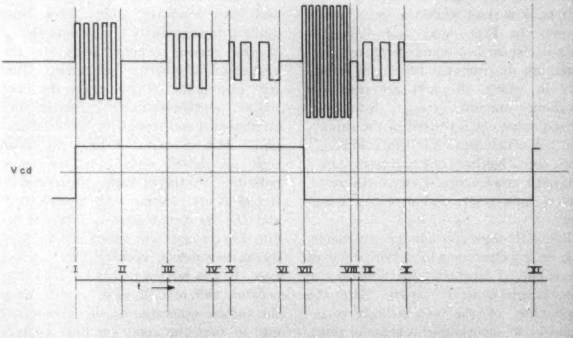

Vcd
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
t

Fig. 7. Typical output frame from the 14-channel system

Fig. 8. Output of sequential set circuit. The square wave is the supply voltage V_{CD}. Arrow shows end of one frame. All cores reset by maximum amount

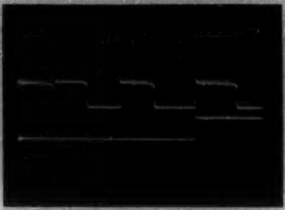


Fig. 9. Output of first 6 stages. Pulses, in order, are from: 1st stage, 3rd, 4th, 5th ballast resistor and stage 6. The signal for core 2 is at its maximum value

Fig. 10. Same as Fig. 9, except that signal for stage 3 is at its maximum value, and core 2 is completely reset

where V_{Tib} is the voltage of the b transducer, and $I_L r_{ib}$ is the voltage drop across the corresponding b rectifier. The rectifiers *CR1b*, *CR3b*, etc. are provided to isolate the inactive stages from the active ones and to block voltages induced in the load windings during reset.

A typical output cycle is shown in Fig. 7. The initial high-frequency burst shown in the interval I to II has its duration controlled by transducer *T1a* and its frequency by *T1b*. Since these two are independent, the burst gives two channels of information. The blank period II to III is controlled by transducer *T2a;* thus, it affords a third channel. (It is perfectly feasible, and may be desirable in certain cases, to connect the output windings of stage *2* through a transducer to the high-frequency Royer. In this case, the wave train would be a succession of bursts of varying duration and frequency following each other so closely as to be practically continuous. Thus, our basic circuit would yield three more channels, for a total of 17.) In the interval III to IV, the high-frequency Royer is on for a period determined by transducer *T3a* at a frequency determined by transducer *T3b*. Thus, in general, the on and off times of the high-frequency Royer are made to convey information. The on times carry further information in the form of the frequency of oscillation.

The interval from VI to VII is the time during which the supply voltage V_{CD} is applied to the ballast resistor R_{Bf}. It may vary in magnitude from zero to practically the full half-cycle of the supply voltage. The interval X to XI is another such period. In this case, it affords a convenient method of "tagging" the cyclically repeating output, since it may be made longer than any other interval.

Since the supply source in itself is a multivibrator, and its frequency, and hence its period, depend on the voltage driving it, a measure of the period gives a measure of the voltage source driving it. Since the output of transducers frequently depends on the voltage source used with them, this information is useful and is counted as another channel. Thus we have a channel for each of the 13 transducers shown in Fig. 10 and an additional channel giving information on the voltage source, or 14 channels in all.

The intermodulation of the outputs of the sequentially set cores was checked by observing the change in one particular output pulse when all the other pulses were changed over their maximum range. A Tektronix *531* Scope was used. It was found that the intermodulation variation was less than the width of the oscilloscope trace for a total pulse width of 1.6 milliseconds. From the time scale and trace used, this is less than 1%.

Figs. 8, 9, and 10 show typical cases of the sequential circuit output. The multiplexed output using the high-frequency multivibrator is not shown, since it originated with Mr. Rochelle, who has discussed it elsewhere.[3]

Discussion

Though the circuit is flexible as to the number of stages which can be employed in each bank, it is difficult to predict what the limiting number might be. One apparent limitation is the resistive losses in the windings of the earlier stages. Use of large wire, few turns, and a high supply voltage would clearly permit the use of many more than the five stages tested. The more stringent limitation probably lies in the frequency limit of core switching with transistor drive. Suppose, for example, that the sampling (or frame) rate is high and the number of stages n is large. Then, since the time available to set a given core is $1/2f$ divided by n, or $1/2nf$, set may have to be accomplished in a microsecond or so, a time too short to be achieved by presently available components. Until this limitation can be overcome, a larger number of stages can be achieved only at the expense of a lower frame rate.

The frame rate is set by the fundamental frequency of the a-c supply. By present techniques, this rate can certainly be increased to a few kilocycles, provided the number of stages is not too large. A limitation here is the frequency capability of the a-c supply transistors, which in turn depends on the power they must handle.

Though the reset circuits shown draw the resetting voltage V_{AB} from a common supply, separate resetting voltage windings can be provided for each stage, thereby permitting either grounded or ungrounded input.

The possibility of using a synchronous decommutating circuit at the receiver end of a telemetering link presents problems. While it should not be hard to arrange for a synchronous multivibrator on the ground, the fact that the output of a given core follows immediately on the termination of the preceding one, which is pulse-width modulated, means that the output of a given channel does not occur at a fixed time in the frame. However, this difficulty can be overcome through the use of a diode-switching matrix and counters.

The output of the present circuit is similar to that of the commutating and encoding system to be used in the earth satellite.[3]

Conclusion

The sequential set circuit is seen to be basically an application of magnetic amplifier circuitry and as such displays the usual characteristics of such circuits. It handles direct voltages with facility, has a rather limited bandwidth, and the output occurs about one half-cycle later than the input. Considerable power gain is also possible, particularly when the internal impedance of the supply source

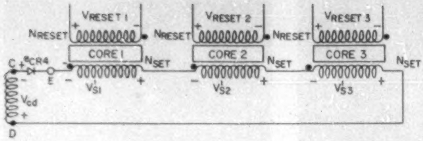


Fig. 11. The voltages in the set circuit during reset

is kept low. (The only inherent power limitation is in the transistors.) In conclusion, it may be stated that the components of the present circuit are reliable; the circuit itself is simple and is capable of being reduced to a small package which can be potted and hence made mechanically very rugged.

Appendix

The conditions which must be satisfied for proper operation of the circuit are stated here mathematically and from them are derived the necessary relationships between supply voltage, resetting voltage, control voltage, etc. This derivation is based on the assumptions that: 1. all stage cores are identical, and, 2. all biasing and signal sources are identical.

The biasing sources, though indicated as separate units, may represent the minimum (nonzero) values of the signal sources, or a combination of this minimum value and padding resistance or separate voltage sources. In this context, the signal voltages V_{ci} are understood to be the voltage in excess of the minimum voltage.

The usual first-order approximations, common in magnetic amplifier design, are made. That is, the magnetizing current I_m is assumed to be constant during a traversal of the hysteresis loop regardless of the voltage applied, and fixed rectifier and resistance drops are assumed.

Condition 1 is illustrated in Fig. 11 and may be stated as: The sum of the voltage induced in the set windings during reset cannot be greater than the supply voltage V_{CD}. V_{ri} is the voltage applied to the reset windings of the *i*th stage, V_{si}' is the voltage induced in the set windings of the *i*th stage, V_{CD} is the supply voltage, *eCR4* is the voltage across the blocked rectifier, N_s is the number of turns in the set windings, and N_r is the number of turns in the reset windings.

In equation form,

$$V_{CD}-(V_{s1}'+V_{s2}'+V_{s3}')=eCR4\geqq 0$$

or, in terms of the resetting voltages,

$$V_{CD}-\frac{N_s}{N_r}(V_{r1}+V_{r2}+V_{r3})=eCR4\geqq 0 \qquad (2)$$

For simplification, the assumption has been made that all set windings have the same number of turns and that all reset windings are identical.

Condition 2 is that the total volt-seconds reset of the cores which takes place during the entire reset half-cycle must be less than or equal to the total volt-second capability of the supply voltage, V_{CD}, during the set half-cycle.

Fig. 12 is an illustration of the resetting circuit of the *i*th stage during reset. Where V_{AB} = reset-source voltage, V_{bi} is the bias voltage, V_{ci} is the control voltage, V_{ri} is the voltage applied to reset windings, and V_{di} is the total voltage drop in rectifier and winding resistance.

By Kirchoff's law,

$$V_{AB}-V_{bi}-V_{ci}-V_{ri}-V_{di}=0$$

from which

$$V_{ri}=V_{AB}-V_{bi}-V_{ci}-V_{di} \qquad (3)$$

Clearly, V_{ri} is the maximum for V_{ci} equal to zero.

The flux change during reset is given by

$$\frac{1}{N_{ri}}\int_0^{T/2} V_{ri}dt=\frac{1}{N_{ri}}\times$$

$$\int_0^{T/2}(V_{AB}-V_{bi}-V_{ci}-V_{di})dt=\Delta\phi \qquad (3A)$$

For complete reset, V_{ci} is zero and

$$\frac{1}{N_{ri}}\int_0^{T/2}(V_{ri})_{\max}dt=\frac{1}{N_{ri}}\times$$

$$\int_0^{T/2}(V_{AB}-V_{bi}-V_{di})dt=2\phi_s \qquad (3B)$$

The cores are set individually as shown in Fig. 13, where V_{si} is the setting voltage, V_{CD} is the supply voltage, and V_t is the lumped transistor and rectifier drop. Summing voltages,

$$V_{CD}-V_{si}-V_t=0$$

$$V_{si}=V_{CD}-V_t$$

The volt-seconds absorbed by this core in the set half-cycle is

$$\frac{1}{N_s}\int_0^{t_1} V_{si}dt=\frac{1}{N_s}\int_0^{t_1}(V_{CD}-V_t)dt$$

Since the cores are read out in sequence, the total volt-seconds absorbed is

$$\frac{1}{N_s}\left[\int_0^{t_1}(V_{CD}-V_t)dt+\int_{t_1}^{t_2}(V_{CD}-V_t)dt+\int_{t_2}^{t_3}(V_{CD}-V_t)dt\right]$$

or

$$\frac{1}{N_s}\int_0^{t_3}(V_{CD}-V_t)dt \qquad (3C)$$

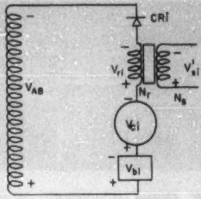


Fig. 12. The voltages in the resetting circuit of the ith stage during reset

It is evident that t_3 must be less than or at most equal to $T/2$; t_3 will have its maximum value when all cores have been reset by a maximum amount. As we have seen, this occurs when all the V_{ci} are zero; hence, using equations 3(B) and 3(C),

$$\frac{1}{N_s}\int_0^{T/2}(V_{CD}-V_t)dt$$

$$=\frac{1}{N_r}\sum_i^n\int_0^{T/2}(V_{AB}-V_{di}-V_{bi})\,dt$$

or

$$\frac{1}{N_s}\int_0^{T/2}(V_{CD}-V_t)dt=3/N_r\times$$

$$\int_0^{T/2}(V_{AB}-V_{di}-V_{bi})dt \qquad (3D)$$

Equation 3 (D) is the mathematical expression of condition 2.

Condition 3 is that the voltage induced in the reset windings during the set half-cycle must not be large enough to unblock the reset circuit rectifier. The resetting circuit and the voltages during set are diagrammed in Fig. 14, where V_{si} is the setting voltage, V_{ri}' is the voltage induced in reset winding during set, V_{ci} is the control voltage, V_{bi} is the bias voltage, e_{ri} is the voltage across blocked rectifier, and V_{AB} is the reset-source voltage. Summing the voltages gives

$$V_{AB}+V_{bi}+V_{ci}-V_{ri}'=eCRi\geqq 0 \qquad (4)$$

which is the equation for condition 3.

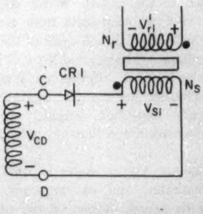


Fig. 13. The voltages in the set circuit of the ith stage during set

Vol.
Pt. 1
195

Thus, the three following equations must be satisfied:

$$V_{CD}-\frac{N_s}{N_r}(V_{r1}+V_{r2}+V_{r3})=eCR4\geqq 0 \qquad (2)$$

$$\frac{1}{N_s}\int_0^{T/2}(V_{CD}-V_t)dt=\frac{3}{N_r}\times$$

$$\int_0^{T/2}(V_{AB}-V_{di}-V_{bi})dt \qquad (3D)$$

and

$$V_{AB}+V_{bi}+V_{ci}-V_{ri}'=eCRi\geqq 0 \qquad (4)$$

We begin by comparing equations 2 and 3(D). By equation 3, and with the assumption that $V_{b1}=V_{b2}=V_{b3}=V_b$ and $V_{d1}=V_{d2}=V_{d3}=V_d$

$$V_{ri}=V_{AB}-V_b-V_{ci}-V_d$$

Before substituting this into equation 2, we note that the condition for equality of equation 3(D) is for the case of maximum reset, i.e., for $V_{ci}=0$.

$$V_{CD}-\frac{N_s}{N_r}[(V_{AB}-V_b-V_d)+(V_{AB}-$$

$$V_b-V_d)+(V_{AB}-V_b-V_d)]=eCR4$$

or

$$V_{CD}-\frac{3N_s}{N_r}(V_{AB}-V_b-V_d)=eCR4 \qquad (5)$$

To reduce equation 3(D), we take advantage of the fact that all the voltages are constant over the time of integration, so we may write

$$\frac{V_{CD}-V_t}{N_s}\,\frac{T}{2}=\frac{3(V_{AB}-V_b-V_d)T}{2N_r}$$

$$(V_{CD}-V_t)=\frac{3N_s}{N_r}(V_{AB}-V_b-V_d)$$

or

$$V_{CD}-\frac{3N_s}{N_r}(V_{AB}-V_b-V_d)=V_t \qquad (6)$$

Comparing equations 5 and 6, we see that they are identical if *eCR4* equals V_t. Since V_t is positive, *eCR4* is greater than zero, and these two conditions are satisfied. (Examining equation 2, we see that if *eCR4* is greater than zero for the maximum values of $V_{r1}+V_{r2}+V_{r3}$, it will always be greater than zero for any value of the sum less than the maximum.)

To compare equation 4 with equations 2 or 3(D), we proceed to put it into an equivalent form, again setting V_c equal to zero. Since equation 4 describes a situation which exists during reset, we note that by transformer action

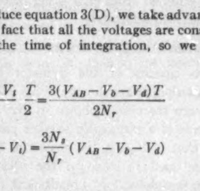


Fig. 14. The voltages in the reset circuit of the ith stage during set

$$\frac{V_{CD}-V_t}{N_s}=\frac{V_{ri}'}{N_r}$$

But by equation 4 $V_{ri}'=V_{AB}+V_b-eCRi$ substituting,

$$\frac{V_{CD}-V_t}{N_s}=\frac{1}{N_r}(V_{AB}+V_b-eCRi)$$

$$V_{CD}=\frac{N_s}{N_r}(V_{AB}+V_b-eCRi)+V_t$$

$$V_{CD}-\frac{N_s}{N_r}(V_{AB}+V_b-eCRi)=V_t \qquad (7)$$

Comparing equation 6 with 7, we see that we must have

$$\frac{3N_s}{N_r}(V_{AB}-V_b-V_d)=\frac{N_s}{N_r}(V_{AB}+V_b-eCRi)$$

Reducing and collecting terms,

$$2V_{AB}=4V_b-eCRi+3V_d$$

or

$$V_b=1/2V_{AB}+1/4eCRi-3/4V_d$$

If we require that the resetting rectifier be blocked by an amount equal to its forward drop (clearly an adequate condition when silicon rectifiers are used),

$$V_d=eCRi \text{ and } V_b=1/2(V_{AB}-V_d) \qquad (8)$$

Thus, the three simultaneous conditions for operation can be met by making V_b one-half of V_{AB} less the resetting circuit drops.

If a padding resistor R_b is to be used instead of the bias source *Vb*, the calculations must be modified slightly. Equation 4 becomes

$$V_{AB}+V_c-V_r'=eCRi\geqq 0 \qquad (4A)$$

since in setting it up we assumed no current flow, i.e., the rectifier blocked. Equations 5 and 6 are changed to the extent that V_b is replaced by I_mR_b, since in these equations magnetizing current flow was assumed. That is,

$$V_{CD}-V_t=\frac{3N_s}{N_r}(V_{AB}-I_mR_b-V_d) \qquad (6A)$$

Reducing equation 4(A) as we did equation 4,

$$\frac{V_{CD}-V_t}{N_s}=\frac{V_{ri}'}{N_r}$$

but $V_{ri}'=V_{AB}-eCRi$, since V_{ci} is zero, or

$$(V_{CD}-V_t)=\frac{N_s}{N_r}(V_{AB}-eCRi) \qquad (7A)$$

Thus we must equate

$$\frac{3N_s}{N_r}(V_{AB}-V_d-I_mR_b)=\frac{N_s}{N_r}(V_{AB}-eCRi)$$

Reducing and collecting,

$$2V_{AB}-3I_mR_b=3V_d-eCRi$$

$$3I_mR_b=2V_{AB}+eCRi-3V_d$$

Recalling that we have made $V_d=eCRi$, the result is

$$R_b=2/3\frac{(V_{AB}-V_d)}{I_m} \qquad (8A)$$

or R_b should be just 2/3 of the ratio of V_{AB} less V_d to I_m

Equations 8 and 8(A) are easily generalized to the case of *n* stages. Equation 8 becomes

$$V_b=\frac{(n-1)}{(n+1)}(V_{AB}-V_d) \qquad (9)$$

and equation 8(A) takes the form

$$R_b=\frac{(n-1)}{n}\,\frac{V_{AB}-V_d}{I_m} \qquad (9A)$$

Using the preceding equations, the explicit forms of the reset equation 3(B) can be written. If voltage bias is used,

$$\frac{1}{N_r}\int_0^{T/2}\left\{V_{AB}-\left[\frac{n-1}{n+1}(V_{AB}-V_d)\right]\right.$$

$$\left.-V_d\right\}dt=2\phi_s$$

and

$$\frac{2}{n+1}\,\frac{1}{N_r}\int_0^{T/2}(V_{AB}-V_d)dt=2\phi_s \qquad (10)$$

If resistance bias is used,

$$\frac{1}{N_r}\int_0^{T/2}\left\{V_{AB}-I_m\left[\frac{(n-1)}{n}\times\right.\right.$$

$$\left.\left.\frac{V_{AB}-V_d}{I_m}\right]-V_d\right\}dt=2\phi_s$$

and

$$\frac{1}{n}\,\frac{1}{N_r}\int_0^{T/2}(V_{AB}-V_d)dt=2\phi_s \qquad (10A)$$

The relation between supply voltage and reset-source voltage is found for the voltage bias case by substituting equation 9 into equation 6.

$$V_{CD}-V_t=n\frac{N_s}{N_r}\left\{V_{AB}-\right.$$

$$\left.\left[\frac{n-1}{n+1}(V_{AB}-V_d)\right]-V_d\right\}$$

$$\frac{V_{CD}-V_t}{N_s}=\frac{n}{n+1}\,2\frac{(V_{AB}-V_d)}{N_r} \qquad (11)$$

For resistance bias, we use equations 9(A) and 6.

$$V_{CD}-V_t=n\,\frac{N_s}{N_r}\times$$

$$\left\{V_{AB}-I_m\left[\frac{n-1}{n}\,\frac{(V_{AB}-V_d)}{I_m}\right]-V_d\right\}$$

$$\frac{V_{CD}-V_t}{N_s}=\frac{V_{AB}-V_d}{N_r} \qquad (11A)$$

To show the relations between signal voltage, reset-source voltage, supply voltage, and core flux, refer to equation 3(B). Maximum reset occurs for $V_{ci}=0$ and V_{ri} a maximum. The other extreme of no reset corresponds to $V_{ri}=0$, and by equation 3 must occur for

$$0=V_{AB}-V_b-V_c-V_d$$

or

$$V_{c\,\max}=V_{AB}-V_b-V_d$$

Putting in the value of V_b from equation 9 we have

$$V_{c\,\max}=V_{AB}-\frac{n-1}{n+1}(V_{AB}-V_d)-V_d$$

$$V_{c\,\max}=\frac{2}{n+1}(V_{AB}-V_d) \qquad (12)$$

A comparison with equation 10 shows that

$$\frac{1}{N_r}\int_0^{T/2} V_{c\,\max}dt=2\phi_s$$

Similarly, in the case for a padding resistor,

$$V_{c\,\max}=\frac{1}{n}(V_{AB}-V_d) \qquad (12A)$$

and

$$\frac{1}{N_r}\int_0^{T/2} V_{c\max}dt=2\phi_s$$

Relating V_c to V_{cb} through equations 11 and 12, we find

$$\frac{V_{CD}-V_t}{N_s}=n\,\frac{V_{c\,\max}}{N_r} \qquad (13)$$

for both bias voltage and padding resistor.

References

1. INVENTIONS WANTED BY THE ARMED FORCES (booklet). Office of Technical Services, National Inventions Council, U. S. Department of Commerce, Washington, D. C., 1957, p. 16, problem 570.

2. THE UTILIZATION OF DOMAIN WALL VISCOSITY IN DATA HANDLING DEVICES, V. L. Newhouse. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, Nov. 1957, p. 1484.

3. CYCLOPS CORES SIMPLIFY EARTH-SATELLITE CIRCUITS, W. Matthews, R. W. Rochelle, C. B. House, R. L. Van Allen, D. H. Shaffer, J. C. Schaffert. *Electronics*, New York, N. Y. vol. 31, Feb. 28, 1958, pp. 56–63.

Linear Power Amplifiers Using Dynistors or Trinistors

FRANK J. HIERHOLZER, JR.
ASSOCIATE MEMBER AIEE

THE DEVELOPMENT of solid-state 2- and 3-terminal switching devices has made possible the design of physically small linear power amplifiers which have a large power-handling capability. Two- and 3-terminal switching devices will be referred to as "dynistors" and "trinistors" throughout most of this paper. Germanium dynistors and trinistors make excellent output devices for the linear power amplifier because: 1. The OFF state impedances are approximately 50,000 ohms; 2. The ON state impedances are approximately 0.05 ohm; 3. The dynistors and trinistors can be operated at much higher current densities than transistors; 4. The dynistors and trinistors need only a pulse to turn on. Once the dynistor or trinistor is on, the device can be turned off by reducing the current through it to a very small value. The characteristics of the germanium dynistor and trinistor are summarized in Appendix I.

The linear power amplifiers described in this paper use dynistors and trinistors as switches in series with the load or output. A magnetic amplifier is used as a transducer to translate input signal magnitude changes into electrical phase-angle changes. The voltage produced by the transducer at the electrical phase angle is impressed across a pulse network which passes the dynistor or trinistor turn-on pulse of current. The dynistors or trinistors turn off just before the current through the load becomes zero each half-cycle of the alternating current. Output waveforms from the dynistor or trinistor linear power amplifiers are very similar to those obtained from a magnetic amplifier.

This paper is limited to describing only two of the many possible magnetic-amplifier transducers and one of the many possible trinistor output circuits of the dynistor and trinistor linear power amplifiers. The output functions for both the dynistor and trinistor linear power amplifiers are defined in Appendix II.

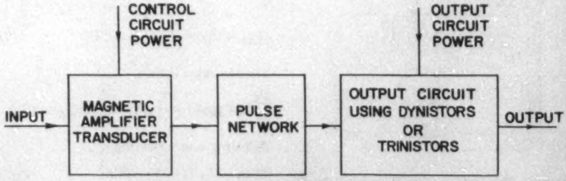
Basic Theory of Operation

Either the dynistor or the trinistor linear power amplifier can be divided into three functional blocks. Fig. 1 shows the three blocks, which are: 1. the magnetic amplifier transducer, 2. the pulse network, and 3. the dynistor or trinistor output circuit. Because the dynistor and trinistor linear power amplifiers are somewhat different from each other, they will be described separately.

DYNISTOR LINEAR POWER AMPLIFIER

Two circuits of the dynistor linear power amplifier are shown in Figs. 2 and 3. Although the circuit shown in Fig. 3 can be made in a physically smaller package than that in Fig. 2, the circuit in Fig. 2 will be used to describe the operation of the amplifier. A breadboard model of the circuit which is shown in Fig. 2 was thoroughly tested at room temperature. Because the dynistors in the output circuit determine the design of the pulse network and the magnetic-amplifier transducer, the output circuit will be described first.

The output circuit consists of two diodes and two dynistors. The diodes and dynistors are arranged so that the output current does not flow through the control circuit. Each half-cycle of the alternating output voltage, one of the



CONTROL CIRCUIT POWER
OUTPUT CIRCUIT POWER
INPUT
MAGNETIC AMPLIFIER TRANSDUCER
PULSE NETWORK
OUTPUT CIRCUIT USING DYNISTORS OR TRINISTORS
OUTPUT

Fig. 1. Block diagram of linear power amplifier

Paper 58-1206, recommended by the AIEE Semiconductor Metallic Rectifiers and the Solid State Devices Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 30, 1958; made available for printing August 15, 1958.

FRANK J. HIERHOLZER, JR., is with the Westinghouse Electric Corporation, Pittsburgh, Pa.

The author wishes to thank J. Rosa from Westinghouse Director Systems and G. Metcalf and A. Zuckerwar from Westinghouse Materials Engineering for their contributions to this paper.

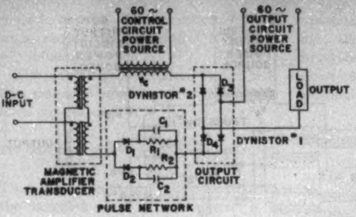


Fig. 2 (left). Dynistor linear power amplifier

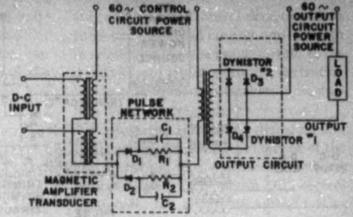


Fig. 3 (right). One possible variation for size reduction of dynistor amplifier

regular diodes shunts the output current by the control circuit. The regular diodes also shunt the control current by the output circuit. The peak value of the alternating output voltage must be less than the breakdown voltage (see Appendix I) of either dynistor. Very little load current (leakage and saturation current of the dynistor) will flow through the dynistors until the alternating control voltage turns on the dynistors. The peak of the alternating control voltage must be greater than the breakdown voltage of each dynistor. The firing angle shown in Fig. 4 is determined by the magnetic amplifier transducer but it has a maximum and minimum value determined by the alternating control voltage. If V_B is the breakdown voltage of each dynistor, then ϕ maximum and ϕ minimum are

$$\phi \text{ max} = \text{arc sin } \frac{V_B}{V_{c \text{ max}}}$$

where ϕ max $> 90°$; and

$$\phi \text{ min} = \text{arc sin } \frac{V_B}{V_{c \text{ max}}}$$

where ϕ min$<$90 degrees.

The percentage of each half-cycle over which the magnetic-amplifier transducer will have control is

$$\text{Per cent of control} = \frac{\phi \text{ max} - \phi \text{ min}}{180°}(100)$$

where ϕ max and ϕ min are firing angles given in degrees.

Because the breakdown voltage of each dynistor is determined by the peak of the alternating output circuit supply voltage, the per-cent control desired would determine the control circuit alternating voltage.

A pulse network is inserted in series with the dynistor output circuit and the magnetic-amplifier transducer. The pulse network must pass the voltage pulse created by the saturation of the magnetic-amplifier transducer. After one of the dynistors turns on or changes from the high-impedance state to the hyperconductive state, the pulse network limits the current in the control circuit. One half of the pulse network controls the control circuit current each half-cycle of the alternating control voltage after one of the magnetic-amplifier cores saturates. The capacitors in the pulse network must be large enough to remain essentially uncharged if one of the magnetic-amplifier cores saturates before the ϕ minimum angle is reached. When either core saturates before ϕ minimum, the dynistor will not turn on and the capacitor charging time constant would be

$$T_{\text{charging}} \approx \frac{V_B}{I_B} C_1 = \frac{V_B}{I_B} C_2$$

The charging time constant must be large in relation to the time of 1/8 cycle of the control voltage. One eighth of a cycle represents an approximation of the time which the control voltage takes to rise from zero to the breakdown voltage of the dynistor

$$T_{\text{charging}} = 10\ (1/480) \text{ seconds}$$

This fixes the minimum value of capacitance. Because the capacitor must be discharged each half-cycle through the resistor in parallel with it, the discharging time constant is

$$T_{\text{discharging}} = R_1C_1 = R_2C_2$$

$$T_{\text{discharging}} \approx (1/2))(1/120) \text{ seconds}$$

To make R_1 and R_2 as large as possible, the minimum value of capacitance should be used. Diodes D_1 and D_2 serve three purposes: 1. They prevent positive feedback from the capacitors to the magnetic amplifier; 2. They allow the capacitors to discharge each half-cycle; 3. They reduce the control circuit current. However, if the diodes D_1 and D_2 block before the end of the gating or output half-cycle of the magnetic amplifier, the linearity of the magnetic amplifier will be affected.

D-c input changes are translated to electrical phase-angle changes with the use of a magnetic-amplifier transducer. The magnetic amplifier must be designed to absorb the total volt-seconds of each half-cycle of the control voltage. Winding resistances of both the transformer and and the magnetic amplifier must be

$$2R_{WMA} + R_{WT} \approx \frac{V_{c \text{ max}} \sin \phi \text{ min} - V_B}{I_B}$$

where

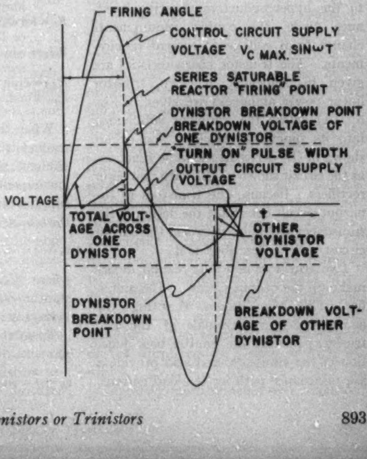


Fig. 4. Waveforms of dynistor linear power amplifier shown in Fig. 2, resistive load

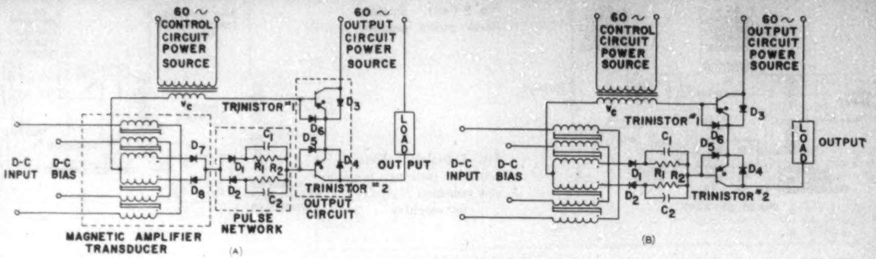


Fig. 5. Trinistor linear power amplifiers

R_{WMA} = winding resistance of the load or gate winding of one core
R_{WT} = winding resistance of the transformer secondary
I_B = dynistor current necessary for voltage breakdown
V_B = dynistor breakdown voltage

Linearity of the dynistor linear power amplifier is largely determined by the linearity of the magnetic amplifier. Because the dynistors have a maximum and minimum firing angle, the dynistor linear power amplifier will not be linear over the full half-cycle. The dynistor nonlinearities occur near the beginning and near the end of each half-cycle where the magnetic amplifiers are nonlinear. For linearity in the middle part of the transfer curve, both the control voltage and the output voltage waveforms should be the same. In other words, if one is a sine wave the other must be a sine wave also. Fig. 4 shows some of the amplifier waveforms near half output for a resistive load.

TRINISTOR LINEAR POWER AMPLIFIER

With the addition of a third electrode to the hyperconductive negative resistance diode, the control requirements change from voltage to current requirements. The trinistor characteristics are given in Appendix I. Two trinistor linear power amplifiers are given in Fig. 5; (A) shows the amplifier subdivided into its functional blocks. The circuit shown in (B) performs the same functions as the circuit shown in (A), with two fewer rectifiers. Because the trinistors in the output circuit determine the design of the transducer and pulse network, the output circuit will be described first.

Two trinistors and four regular diodes make up the output circuit of the amplifier. The two diodes shunting the emitter-collector terminals of the two trinistors protect the emitter-base junction of the trinistors, and also provide a low-impedance path for the load current. The two diodes shunting the base-emitter junction of each trinistor protect each emitter-base junction and provide a low-impedance path for control current. When base current equal to I_{b_s} flows in either trinistor, the breakdown voltage (see Appendix I) of the trinistor drops to zero. The maximum and minimum firing angles (see Fig. 6) can be determined by first plotting breakdown voltage versus base current and then obtaining an equation of breakdown voltage as a function of base current for the low breakdown voltage portion of the curve. For the trinistor this equation is

$$v_B = m i_{\text{base}} + C$$

where m and C are constants for a given trinistor.

The base current is

$$i_{\text{base}} = \frac{V_{c\,\max} \sin \phi}{R_{WMA} + 2R_R + R_{WT} + R_{EB}}$$

where

$V_{c\max} \sin \phi = V_{c\max} \sin \omega t$ = instantaneous value of control circuit voltage
R_{WMA} = winding resistance of magnetic-amplifier load winding
R_R = forward resistance of either D_1, D_2, D_5, or D_6 (assuming they are equal)
R_{WT} = winding resistance of transformer secondary
R_{EB} = emitter-base resistance of trinistor
i_{base} = instantaneous value of base current

When the output circuit supply voltage reaches the breakdown voltage of the trinistor, the trinistor will become a very low impedance. The equation for output supply voltage across the trinistor before the trinistor turns on is

$$v_0 = v_T = V_{0\,\max} \sin \phi$$

where $V_{\max} \sin \phi = V_{0\max} \sin \omega t$ = instantaneous value of the output source voltage across one trinistor.

To solve for ϕ minimum, equate v_T and v_B:

$$v_B = v_T = m i_{\text{base}} + C = V_{0\,\max} \sin \phi_{\min}$$

Substituting i_{base} in this equation yields

$$m\left(\frac{V_{c\,\max} \sin \phi_{\min}}{R_{WMA} + 2R_R + R_{WT} + R_{EB}}\right) + C = V_{0\,\max} \sin \phi_{\min}$$

$$\left(\frac{m V_{c\,\max}}{R_{WMA} + 2R_R + R_{WT} + R_{EB}} - V_{0\,\max}\right) \times \sin \phi_{\min} = -C$$

$$\phi \min = \text{arc} \sin \left(\frac{-C}{\dfrac{m V_{c\,\max}}{R_{WMA} + 2R_R + R_{WT} + R_{EB}} - V_{0\,\max}} \right)$$

Note that m is a negative quantity: ϕ minimum<90 degrees.

The equation for ϕ minimum will hold only if the capacitance in the pulse network is large enough so that the base current is determined by the control circuit resistance and instantaneous control voltage. This equation is also derived on the assumption that the cores of the magnetic amplifiers saturate before ϕ minimum and the inductive reactance of the cores after the cores saturate is negligible in comparison to the resistance in the

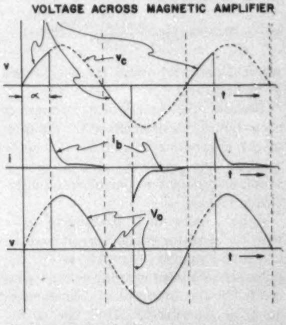


Fig. 6. Trinistor linear power amplifier waveforms, resistive load

control circuit. The maximum firing angle is equal to ϕ maximum $= 180$ degrees $-\phi$ minimum.

The percentage of each half-cycle which the magnetic-amplifier transducer will have control over is

$$\text{Per cent control} = \frac{\phi \max - \phi \min}{180°}(100)$$

Thus the per-cent control obtained with the magnetic-amplifier transducer is a function of the output voltage as well as of the control voltage and the control circuit resistance.

The pulse circuit in the trinistor linear power amplifier must be designed so that 1. the capacitors discharge during one half-cycle of the alternating control voltage; and 2. the charging current of the capacitors is greater than I_{bs} when the magnetic amplifier transducer saturates before ϕ minimum. Both the dynistor and trinistor linear power amplifiers have the same discharge time constants: $T_{\text{discharging}} = R_1C_1 = R_2C_2 \approx 1/2(1/120)$ seconds.

To make R_1 and R_2 as large as possible, the minimum value of capacitance should be used. This capacitance can be calculated using the following relationship:

$$Q = CV$$

$$\frac{\Delta Q}{\Delta t} = C\left(\frac{\Delta V}{\Delta t}\right) = i$$

where $\Delta V/\Delta t$ is the rate of change of the voltage with respect to time near ϕ minimum.

Using a straight line approximation for $\Delta V/\Delta t$ at small angles gives

$$\frac{\Delta V}{\Delta t} = \frac{V_{c\,\max}}{1/240}$$

$$\frac{\Delta V}{\Delta t} = 240\ V_{c\,\max}$$

Substituting $\Delta V/\Delta t$ in the equation for i gives

$$i \approx C(240\ V_{c\,\max})$$

but $i = I_{bs}$ if the trinistors are to turn on at ϕ minimum. Therefore the capacitance C_1 and C_2 is

$$C_1 = C_2 \approx \frac{I_{bs}}{240\ V_{b\,\max}}$$

The magnetic-amplifier transducer translates d-c input changes into electrical phase-angle changes. The discussion of the transducer in the section "Dynistor Linear Power Amplifier" applies to the trinistor linear power amplifier, except for the remarks concerning the winding resistance. The physical size of the magnetic-amplifier transducer is the limiting factor for the control voltage and load-winding resistance. The higher the control voltage, the larger the magnetic amplifier must be. Load-winding resistance can be decreased by increasing the size of the magnetic amplifier.

Discussion

The trinistor and the dynistor linear power amplifiers will be compared in this section. Linearity of both amplifiers with inductive and capacitive loads as well as resistive loads will be discussed. Although neither amplifier has been tested in ambient temperatures above and below normal room temperature, some temperature limitations will be given.

In both amplifiers the load or output current must be independent of the control current. This enables the control circuit of the amplifier to function independently of the output. Large power gains between the control circuit and the output circuit of the dynistors and trinistors are then possible. The actual power gain between the control circuit and the output circuit of the dynistor amplifier is approximately 20 db (decibels). This is less than the 45-db gain obtained in the trinistor amplifier because of the high voltage in the dynistor control circuit. Breadboard models of the dynistor amplifier had over-all power gains of approximately 45 db. Breadboard models of the trinistor amplifier had over-all power gains greater than 60 db. No attempt was made to optimize the gain of either amplifier.

High control circuit voltage in the dynistor amplifier also increases its physical size. The trinistor amplifier is about one tenth the size of the dynistor amplifier. The transformer and magnetic-amplifier transducer in the dynistor amplifier are much larger than those in the trinistor amplifier.

Both the amplifiers function best when the maximum output power is between 500 and 1,000 watts. This 1,000-watt limitation of the breadboard models is not an inherent limitation of the amplifier. One thousand watts is approximately the maximum power-handling capability of present dynistors and trinistors without forced cooling. When dynistors and trinistors which control more power are developed. the amplifiers will be able to control more power. At these output powers the load current is much greater that the sustaining current of the dynistors and trinistors. Because the sustaining currents of the dynistors and trinistors are low in relation to the load or output current, the dynistors and trinistors remain in the hyperconductive state after the pulse from the control circuiting turns them on. The load current must be high in relation to the sustaining current of the dynistors and trinistors for the derivation of the maximum and minimum conduction angles to be valid.

Because the germanium dynistors and trinistors used in the output circuit of the amplifiers will not operate in ambient temperatures much above 75 degrees centigrade, the maximum ambient temperature of these amplifiers would be less than 75 degrees centigrade. The minimum ambient temperature would be determined by the nonlinearity of the magnetic - amplifier transducer. The change of resistance of the copper windings would have to be compensated for by the use of a negative temperature coefficient thermistor in the magnetic-amplifier input winding circuit.

The magnetic-amplifier transducer used in the trinistor linear power amplifier is a better transducer than the series saturable reactor used in the dynistor linear power amplifier. With the use of the steady-state output functions given in Fig. 7 for the two transducers, a comparison of the two can be made. The firing angle of the dynistor linear power amplifier is dependent upon the impedance in the gate or load circuit of the magnetic amplifier. The pulse network in the gate circuit of the magnetic amplifier is a nonlinear impedance with will introduce nonlinearities into the transfer function of the dynistor amplifier. Magnetic-amplifier load resistance does not affect the firing angle of the transducer used in the trinistor amplifier.

Load power factor does influence the operation and linearity of either amplifier. If the load is capacitive and the time constant of the capacitor-charging circuit is much less than the period of 1/2 cycle of the output circuit supply voltage, the capacitor will charge up and the dynistor or trinistor will change to the high-impedance state. The output power will not be a linear function of the input power. With inductive and resistive loads, the amplifiers will have an output power which is a linear function of the input power. With inductive loads, output current continues to flow after the output circuit supply voltage passes through zero (the circuit becomes free wheeling). The dynistors and trinistors do not transfer to the high-impedance state until after the load current becomes less than the sustaining current of the dynistors and trinistors. The resistive load operation of the amplifiers is shown in Figs. 4 and 6.

Linearity measurements were taken on

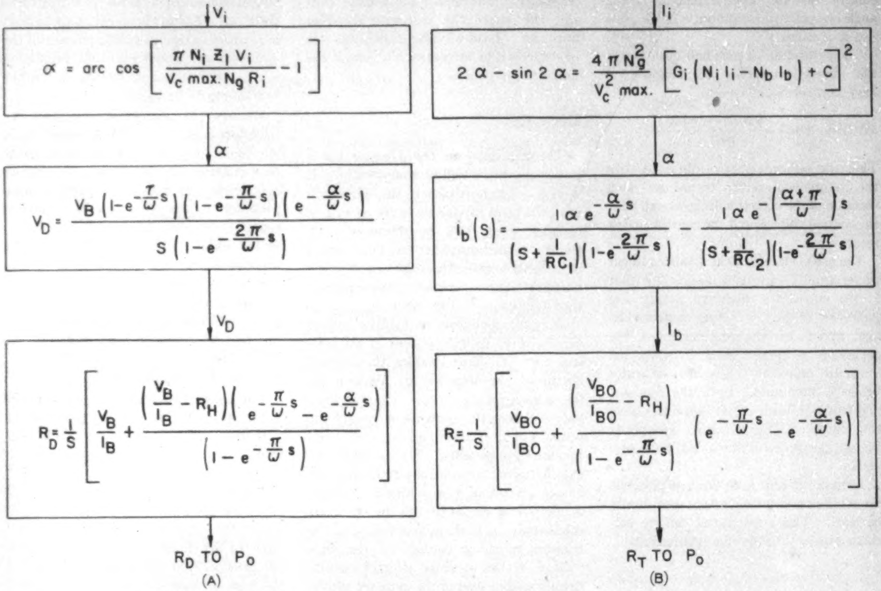

V_i
$\alpha = \text{arc cos}\left[\frac{\pi N_i Z_1 V_i}{V_c\ \text{max.}\ N_g R_i} - 1\right]$
α
$V_D = \frac{V_B\left(1-e^{-\frac{\tau}{\omega}s}\right)\left(1-e^{-\frac{\pi}{\omega}s}\right)\left(e^{-\frac{\alpha}{\omega}s}\right)}{s\left(1-e^{-\frac{2\pi}{\omega}s}\right)}$
V_D
$R_D = \frac{1}{S}\left[\frac{V_B}{I_B} + \frac{\left(\frac{V_B}{I_B} - R_H\right)\left(e^{-\frac{\pi}{\omega}s} - e^{-\frac{\alpha}{\omega}s}\right)}{\left(1-e^{-\frac{\pi}{\omega}s}\right)}\right]$
R_D TO P_o
(A)
I_i
$2\alpha - \sin 2\alpha = \frac{4\pi N_g^2}{V_c^2\ \text{max.}}\left[G_i\left(N_i I_i - N_b I_b\right) + C\right]^2$
α
$i_b(s) = \frac{I\alpha e^{-\frac{\alpha}{\omega}s}}{\left(S+\frac{1}{RC_1}\right)\left(1-e^{-\frac{2\pi}{\omega}s}\right)} - \frac{I\alpha e^{-\left(\frac{\alpha+\pi}{\omega}\right)s}}{\left(S+\frac{1}{RC_2}\right)\left(1-e^{-\frac{2\pi}{\omega}s}\right)}$
I_b
$R_T = \frac{1}{S}\left[\frac{V_{BO}}{I_{BO}} + \frac{\left(\frac{V_{BO}}{I_{BO}} - R_H\right)}{\left(1-e^{-\frac{\pi}{\omega}s}\right)}\left(e^{-\frac{\pi}{\omega}s} - e^{-\frac{\alpha}{\omega}s}\right)\right]$
R_T TO P_o
(B)

Fig. 7. Block diagrams of power amplifiers showing output functions of each block, steady state conditions

A—Dynistor power amplifier
B—Trinistor power amplifier

breadboard models of the dynistor and trinistor amplifiers for both inductive and resistive loads. In the region of output powers from 10% to 90% of maximum output power, the dynistor linear power amplifier had a variation in linearity of ±6%, and the trinistor amplifier of approximately ±3%. The maximum output power of amplifiers when the linearity tests were made was 500 watts. No difference in linearity was noticed when the load was changed from a resistive to an inductive load with a power factor of 60% with either amplifier. The larger deviation from linearity of the dynistor linear power amplifier shows the nonlinearity introduced by the impedance change of the pulse network. No attempt was made to obtain the best possible linearity from either amplifier. Negative feedback from the output to the magnetic-amplifier transducer would improve linearity.

Conclusions

1. Line voltage variations will not affect the linearity of either amplifier appreciably. Negative feedback may be necessary to minimize changes in output when the line voltage of the trinistor linear power amplifier is changed.

2. Linearity of the two amplifiers is primarily dependent upon the linearity of the magnetic-amplifier transducer.

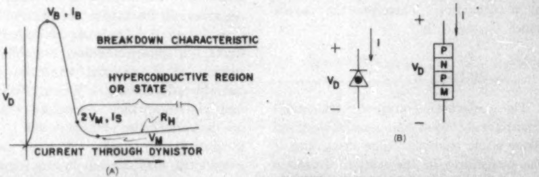
3. Both amplifiers perform best with the smallest permissible load impedance.

4. The trinistor linear power amplifier can be made physically smaller than the dynistor linear power amplifier.

5. Both amplifiers are much smaller than magnetic amplifiers with comparable maximum output powers.

6. Maximum linearity will be obtained with the use of magnetic-amplifier transducers similar to the one used in the trinistor linear power amplifier. The firing angle of the magnetic amplifier should not be a function of the impedance in the gating circuit of the magnetic amplifier.

7. Both amplifiers will have an output power which is a linear function of input power for both resistive and inductive loads.


V_B, I_B
BREAKDOWN CHARACTERISTIC
HYPERCONDUCTIVE REGION OR STATE
V_D
$2V_M, I_S$
R_H
V_M
CURRENT THROUGH DYNISTOR
(A)
+
I
V_D
−
+
I
P
N
P
M
V_D
−
(B)

V_B = DYNISTOR BREAKDOWN VOLTAGE
I_B = BREAKDOWN CURRENT
I_S = MINIMUM CURRENT WHICH WILL KEEP THE DYNISTOR IN THE HYPERCONDUCTIVE REGION
V_M = MINIMUM VOLTAGE DROP ACROSS THE DYNISTOR IN THE HYPERCONDUCTIVE REGION
R_H = DYNAMIC RESISTANCE OF THE DYNISTOR IN THE HYPERCONDUCTIVE REGION
V_D = VOLTAGE ACROSS DYNISTOR

Fig. 8. A—Dynistor characteristics. B—Dynistor configuration. Note: Polarities shown are necessary for obtaining breakdown characteristic

Appendix I. Summary of Dynistor and Trinistor Characteristics

Dynistor Characteristics

The dynistor breakdown volt-ampere characteristic curve is shown in Fig. 8(A). Circuit symbols used to represent the dynistor with the polarities for the breakdown characteristic are shown in Fig. 8(B); the symbol to the right is not generally used in circuits but it is included to show the physical construction of the dynistor.[1]

To change from the high impedance or OFF state to the hyperconductive or ON state, the dynistor must have a voltage V_B across it, and a current I_B through it. Once the dynistor changes to the hyperconductive state, the dynamic resistance R_H of the dynistor is less than 0.1 ohm. A current through the dynistor equal to or greater than the sustaining current I_s must be present if the dynistor is to remain in the hyperconductive state.

Trinistor Characteristics

Fig. 9(A) shows the volt-ampere breakdown characteristic of the trinistor for various base currents. Two circuit symbols for trinistors are shown in Fig. 9(B); the symbol to the right gives a simplified picture of the physical construction of the trinistor.[1]

The base electrode of the trinistor provides a low-resistance path for base current which reduces the breakdown voltage V_B of the trinistor. When the base current is equal to I_{bs}, the breakdown voltage of the trinistor is reduced to zero. The trinistor will be in the hyperconductive or low-impedance state as long as a base current equal to I_{bs} flows. If the base current is reduced to zero, a current equal to or greater than the sustaining current I_s must flow in the collector circuit if the trinistor is to remain in the hyperconductive state.

Appendix II. Explanation of Output-Input Functions for Amplifiers

The output functions of the three functional blocks of each amplifier are given in Fig. 7. Trinistor and dynistor circuit outputs are given as the change in resistance with occurs when the magnetic-amplifier transducer fires. This change in resistance occurs if the magnitude of the pulses from the pulse network is large enough to turn on the dynistors or trinistors in the output circuit. Because these pulses will cause the dynistors or trinistors to change state, the output functions of both the dynistor and the trinistor output circuits are written as a function of the magnetic-amplifier transducer firing angle.

All output functions given represent steady-state conditions. The transient response of either the dynistor or the trinistor linear power amplifier is governed by the response of the magnetic-amplifier transducer. A low-gain fast-response magnetic amplifier can be used as a transducer to obtain fastest response. The equations will not be derived here, but each will be restated, the symbols explained.

Dynistor Linear Power Amplifier

MAGNETIC-AMPLIFIER TRANSDUCER

$\alpha =$ arc cos

$$\left(\frac{\pi N_i Z_L V_i}{V_{c\max}\, N_g R_i}-1\right)$$

$\alpha =$ firing angle of transducer
$V_i =$ input voltage to transducer
$Z_L =$ effective load impedance of magnetic amplifier
$N_i =$ number of input windings on each magnetic-amplifier core
$V_{c\max} =$ maximum control circuit voltage (magnetic-amplifier load circuit)
$N_g =$ number of gate or load windings on each core
$R_i =$ resistance in series with V_i and input windings of magnetic amplifier

This equation was derived assuming voltage reset of the magnetic amplifier.

PULSE NETWORK

$$V_D=\frac{V_B\left(1-\epsilon^{-\frac{\tau}{\omega}S}\right)\left(1-\epsilon^{-\frac{\pi}{\omega}S}\right)\epsilon^{-\frac{\pi}{\omega}S}}{S\left(1-\epsilon^{-\frac{2\pi}{\omega}S}\right)}$$

where

$V_D =$ control circuit voltage across the dynistors
$V_B =$ breakdown voltage of each dynistor
$\tau/\omega =$ turn-on time of each dynistor
$S =$ Laplace transform variable
$\omega =$ angular frequency of either control voltage or output voltage
$\alpha =$ firing angle of magnetic-amplifier transducer

DYNISTOR OUTPUT CIRCUIT

$$R_D=\frac{1}{S}\left[\frac{V_B}{I_B}+\left(\frac{V_B}{I_B}-R_H\right)\left(\frac{\epsilon^{-\frac{\pi}{\omega}S}}{1-\epsilon^{-\frac{\pi}{\omega}S}}\right)-\left(\frac{V_B}{I_B}-R_H\right)\left(\frac{\epsilon^{-\frac{\alpha}{\omega}S}}{1-\epsilon^{-\frac{\pi}{\omega}S}}\right)\right]$$

where

$R_D =$ instantaneous value of dynistor resistance in series with output or load
$V_B =$ breakdown voltage of dynistors
$I_B =$ breakdown current of dynistors
$R_H =$ dynamic resistance of dynistor in hyperconductive state
$\omega =$ angular frequency of control frequency
$\alpha =$ firing angle of magnetic-amplifier transducer
$S =$ Laplace transform variable

The equation for R_D neglects the minimum drop V_m across the dynistors in the hyperconductive state.

Trinistor Linear Power Amplifier

MAGNETIC-AMPLIFIER TRANSDUCER

$$2\alpha-\sin 2\alpha=\frac{4\pi N_g^2}{V_{c\max}^2}\,[G_i(N_iI_i-N_bI_b)+C]^2$$

where

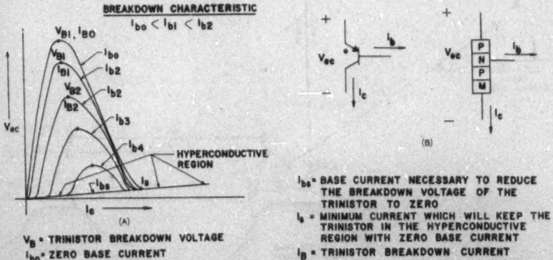
$\alpha =$ firing angle of magnetic-amplifier transducer
$N_g =$ number of gate or load windings on each core
$V_{c\max} =$ maximum control circuit voltage
$G_i =$ gain of core as determined empirically by Roberts test[3]
$C =$ constant determined by Roberts test
$N_i =$ number of input windings on each magnetic-amplifier core
$I_i =$ input current to magnetic-amplifier transducer
$N_b =$ number of bias turns on each magnetic-amplifier core
$I_b =$ bias current to magnetic-amplifier transducer

The equation for α was derived assuming constant magnetomotive force reset using an empirical equation containing contants which are evaluated using Roberts test values.[3]

PULSE NETWORK

$$I_b(S)=\frac{I_\alpha\epsilon^{-\frac{\alpha}{\omega}S}}{\left(S+\frac{1}{RC_1}\right)\left(1-\epsilon^{-\frac{2\pi}{\omega}S}\right)}-\frac{I_\alpha\epsilon^{-\left(\frac{\pi+\alpha}{\omega}\right)S}}{\left(S+\frac{1}{RC_2}\right)\left(1-\epsilon^{-\frac{2\pi}{\omega}S}\right)}$$

where



V_B = TRINISTOR BREAKDOWN VOLTAGE
I_{bo} = ZERO BASE CURRENT

I_{bs} = BASE CURRENT NECESSARY TO REDUCE THE BREAKDOWN VOLTAGE OF THE TRINISTOR TO ZERO
I_s = MINIMUM CURRENT WHICH WILL KEEP THE TRINISTOR IN THE HYPERCONDUCTIVE REGION WITH ZERO BASE CURRENT
I_B = TRINISTOR BREAKDOWN CURRENT

Fig. 9. A—Trinistor characteristics. B—Trinistor configuration: see note for Fig. 8(B)

I_α = peak value of pulse current
$I_\alpha = (V_{c\max} \sin \alpha)/R$
$R = R_{WMA} + 2R_R + R_{WT} + R_{EB}$

TRINISTOR OUTPUT CIRCUIT

$$R_T = \frac{1}{S}\left[\frac{V_{B0}}{I_{B0}} + \left(\frac{V_{B0}}{I_{B0}} - R_H\right)\left(\frac{\epsilon^{-\frac{\pi}{\omega}s}}{1-\epsilon^{-\frac{\pi}{\omega}s}}\right) - \left(\frac{V_{B0}}{I_{B0}} - R_H\right)\left(\frac{\epsilon^{-\frac{\alpha}{\omega}s}}{1-\epsilon^{-\frac{\pi}{\omega}s}}\right)\right]$$

where

V_{B0} = breakdown voltage of trinistor at zero base current
I_{B0} = breakdown current of trinistor at zero base current
R_H = hyperconductive state resistance of trinistor
R_T = instantaneous value of trinistor resistance in series with output or load

Over-all Output Functions

The over-all output-input functions are derived, assuming

$$R_H \ll R_L$$

$$\frac{I_B}{V_B} \gg R_L$$

$$\frac{V_{B0}}{I_{B0}} \gg R_L$$

DYNISTOR LINEAR POWER AMPLIFIER

$$P_0 = \frac{V_{0\max}^2}{2\pi R_0}\left(\pi - \alpha + \frac{\sin 2\alpha}{2}\right)$$

where

$$\alpha = \text{arc cos}\left(\frac{\pi N_t Z_L V_t}{V_{c\max}\ N_g R_t} - 1\right)$$

$V_{\max}$ = maximum voltage of output circuit power source
R_o = output or load resistance
P_o = output power

TRINISTOR LINEAR POWER AMPLIFIER

$$P_0 = \frac{V_{0\max}^2}{4\pi R_0}\left[2\pi - (2\alpha - \sin 2\alpha)\right]$$

where

$$2\alpha - \sin 2\alpha = \frac{4\pi N_g^2}{V^2_{c\max}}\left[G_t(N_t I_t - N_b I_b) + C\right]^2$$

References

1. GERMANIUM POWER SWITCHING DEVICES, J. Philips, H. C. Chang. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, New York, N. Y., vol. ED-5, Jan. 1958, pp. 13–18.

2. HIPERMAG CORES. *Descriptive Bulletin no. 44-750*, Specialty Transformer Department, Transformer Division, Westinghouse Electric Corporation, Greenville, Pa., Aug. 1956, pp. 2–7.

The Expansion of the Pacific Coast Microwave Network

R. G. KUCK
NONMEMBER AIEE

THE PACIFIC COAST Microwave Development during the early formative years could hardly be considered a network in its own right. Although a relatively large use of microwave was planned at this time, the planning was limited to short systems with unquestionable economic advantages over competing facilities.

The initial microwave system installed in the United States for common carrier use provided eight voice channels between Santa Catalina Island, off the coast of Southern California, and the Mutual toll office in Los Angeles. Subsequently, a similar system was employed between Marysville and Redding, Calif. This early equipment was of Western Electric Company manufacture and carried the military designation of *AN/TRC-6*. It operated in the 4,500 to 4,800-mc (megacycle) band utilizing pulse position modulation. Originally designed for 8-channel service, some of these systems were later modified to provide 16 channels. Equipments of this type were continuously employed throughout the Pacific region from 1946 to 1954, and studies conducted on these systems helped pave the way for the future development of microwave communications in the area.

Following a testing program undertaken in the summer of 1946, the first installation of microwave for television transmission took place between Hollywood and Mt. Wilson, Calif., in 1947. Western Electric *TE-1* equipment operating in the then recently assigned allocation of 3,700 to 4,200 mc was installed to interconnect various Hollywood television studios and their transmitters atop 6,000-foot-high Mt. Wilson.

Early in 1948, a new series of tests was undertaken to determine the feasibility of broad-band multiple-section transmission. A number of *TE-1* microwave equipments were arranged as shown in Fig. 1. Both *L* carrier telephone multiplex and video test pattern signals were transmitted over this arrangement. One

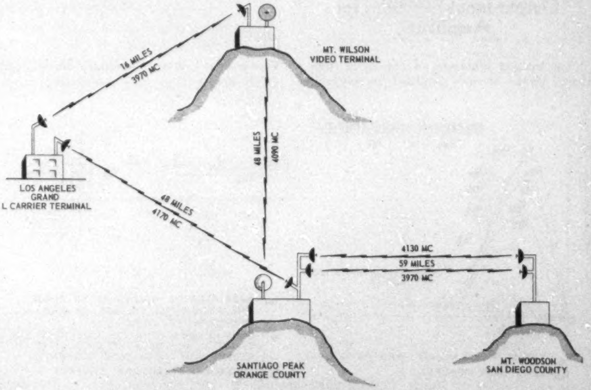


Fig. 1. Los Angeles Grand–Mt. Woodson route

Paper 58-1227, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 21, 1958; made available for printing August 28, 1958.

R. G. KUCK is with the Pacific Telephone and Telegraph Company, Los Angeles, Calif.

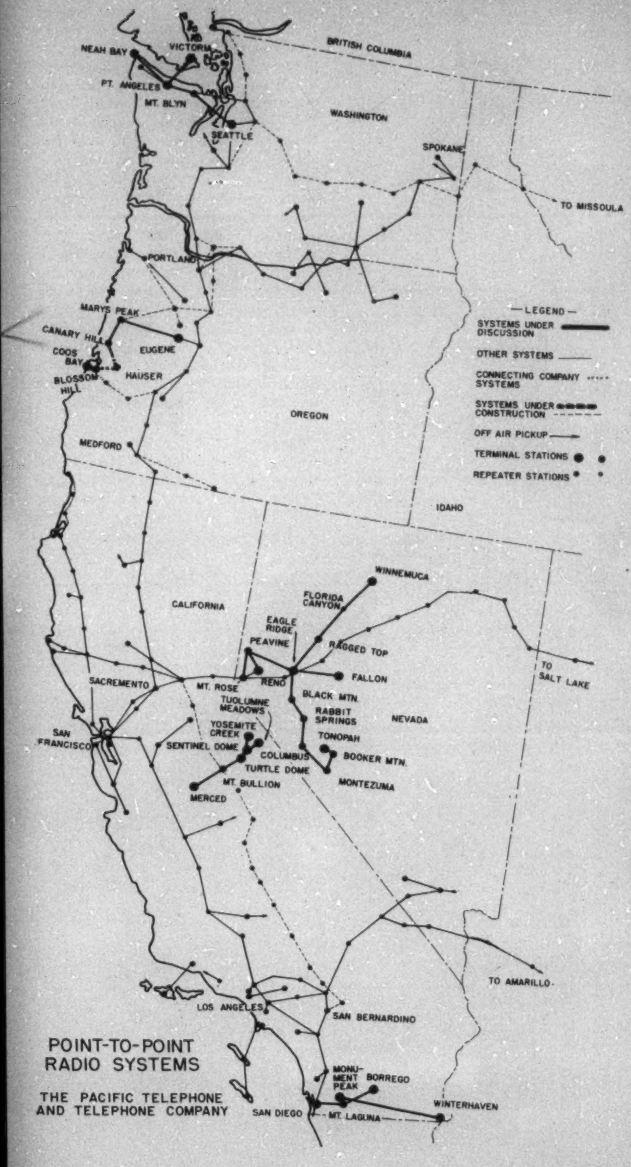

BRITISH COLUMBIA
NEAH BAY
VICTORIA
PT. ANGELES
MT. BLYN
SEATTLE
WASHINGTON
SPOKANE
TO MISSOULA
PORTLAND
MARYS PEAK
CANARY HILL
COOS BAY
BLOSSOM HILL
EUGENE
HAUSER
OREGON
MEDFORD
IDAHO
— LEGEND —
SYSTEMS UNDER DISCUSSION
OTHER SYSTEMS
CONNECTING COMPANY SYSTEMS
SYSTEMS UNDER CONSTRUCTION
OFF AIR PICKUP
TERMINAL STATIONS
REPEATER STATIONS
WINNEMUCA
FLORIDA CANYON
EAGLE RIDGE
PEAVINE
RAGGED TOP
CALIFORNIA
RENO
FALLON
MT. ROSE
SACRAMENTO
TO SALT LAKE
BLACK MTN.
TUOLUMNE MEADOWS
RABBIT SPRINGS
NEVADA
YOSEMITE CREEK
TONOPAH
SAN FRANCISCO
SENTINEL DOME
COLUMBUS
BOOKER MTN.
TURTLE DOME
MONTEZUMA
MT. BULLION
MERCED
TO AMARILLO
LOS ANGELES
SAN BERNARDINO
POINT-TO-POINT RADIO SYSTEMS
THE PACIFIC TELEPHONE AND TELEPHONE COMPANY
MONUMENT PEAK
BORREGO
SAN DIEGO
MT. LAGUNA
WINTERHAVEN

Fig. 2. Pacific Coast microwave system

of these tests involved an *A-B* comparison between long-distance circuits routed directly over an *L* coaxial cable carrier system and a second arrangement connecting the two facilities, microwave and coaxial in tandem. Numerous competent observers talked over these circuits under various traffic-loading conditions. The results were considered highly satisfactory.

Video tests were also conducted. A test pattern furnished directly from a pattern generator was fed through the 3-hop system. (The Mt. Woodson repeater was excluded from this test.) In an *A-B* presentation, the viewers were unable to detect any differences between the original pattern and that same pattern after transmission over the system.

These early installations and test programs paved the way for the broad-band developments to follow.

Later in 1948, an extensive investigation was initated, looking toward the establishment of a Los Angeles–San Francisco microwave route. This route was completed with nine repeater points between the Los Angeles and East Bay Hills areas. The repeater sections range from approximately 16 to 65 miles in length. These longer sections were in order of an experiment, for the nominal sections in use at this time rarely exceeded 35 miles.

As is so often the case, major routes are preceded by temporary or minor route installations. A case in point occurred in 1952. Early in the year, the company received an inquiry requesting television network facilities for a television station in Seattle, Wash. At this time, the video network had been extended from Sacramento, Calif., to Portland, Oreg., by means of an *L1* coaxial cable system, but no further. Although future plans called for a *TD-2* microwave system extension between Portland and Seattle, the normal installation period would have precluded the use of this system to meet the fall of 1952 request. This request was fulfilled by rushing the repeater site buildings to completion and collecting Western Electric *TE* type of equipment from throughout the Bell System. The *TE* equipment provided a temporary service until the originally planned *TD-2* equipment could be installed.

Recent Developments

The so-called major route development, once started, increased at a rapid pace to the extent that facilities are now operating from San Diego, Calif., on the south to Angeles Point, Wash., in the

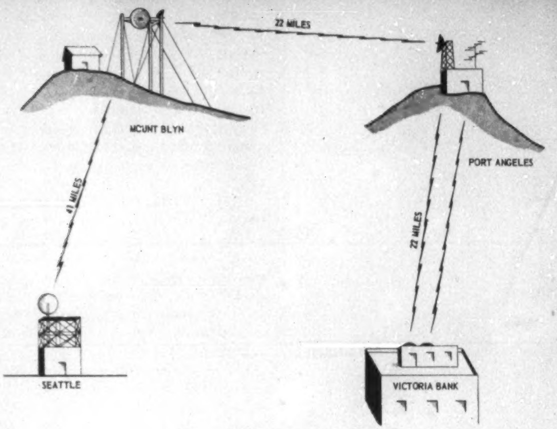


Fig. 3. Seattle–Victoria system

north. In addition, two major microwave routes now interconnect the Pacific Coast and the states in the Middle West and the East. They are the central route from the Bay area through Sacramento and east, and the southern route to Amarillo, Tex., from Los Angeles; see Fig. 2.

Paralleling the growth of long-haul major route facilities, the short-haul systems have also expanded at a rapid rate. These shorter systems operate mainly within the frequencies of 890 to 940, 3,700 to 4,200, and 5,925 to 6,425 mc.

Systems in this category employ a wide variety of equipments of various manufacturers. The original *AN/TRC-6* equipment has now been replaced with Radio Engineering Laboratories' 900-mc equipment and the islands of San Clemente and Santa Rosa have been linked to the mainland in like manner.

Numerous interesting solutions to difficult problems have been realized through the use of short-haul microwave systems. The following examples are more or less typical of the systems installed throughout the area.

In the far northern reaches of the state of Washington is an area known as the Olympic Peninsula. This area is generally primitive and therefore ideally suited for radio development. A microwave system was engineered to provide not only telephone growth to this remote area but to also supplant the then existing Seattle, Everett, Bellingham, Victoria cable, submarine cable, and open-wire composite route. The submarine portion of this latter facility had required almost continual repair because of armor wire abrasions.

On July 14, 1953, a 3-section system utilizing Lenkurt *72A* equipment with Lenkurt *33C* channelizing was placed in service; see Fig 3. The first two sections north of Seattle operate in the 900-mc band and the last section between Port Angeles, Wash., and Victoria, B. C., operates in the 450-mc band. The radio equipment is provided in full duplicate.

Based upon the experience gained from this installation and others in the area, an emergency radio system was placed in operation on December 1, 1957. This new system operates between Angeles Point and an Air Force installation at Neah Bay. It is interesting to note that, although the path is 55 miles in length and "over water," no stand-by or diversity arrangements were employed. The system provides eight circuits to the Neah Bay site in addition to those furnished by an open-wire and a buried cable route. This latter route parallels the highway serving the Neah Bay location; and unfortunately when earth slides disrupt the communication facilities, they also invariably block the highway, thereby complicating the facility restoration. This emergency system has removed the critical time element associated with plant repair while also improving the reliability of the circuits. These are examples where, by the use of radio, the communication services to an area of difficult accessibility have been materially improved.

The State of Oregon, like most of the area in the northwestern portion of the country, is divided by mountain ranges lying in a north-to-south direction. The population growth in the state has taken place mainly along the coast and in the valleys between these ranges. As a means of furnishing relief to the coastal area in the vicinity of Coos Bay, the West Coast Telephone Company and the Pacific Telephone and Telegraph Company jointly implemented a microwave system referred to as the Coos Bay–Eugene system; see Fig. 4. This combined effort was engineered with the use of three intermediate repeater locations designated Hauser, Canary Hill, and Mary's Peak.

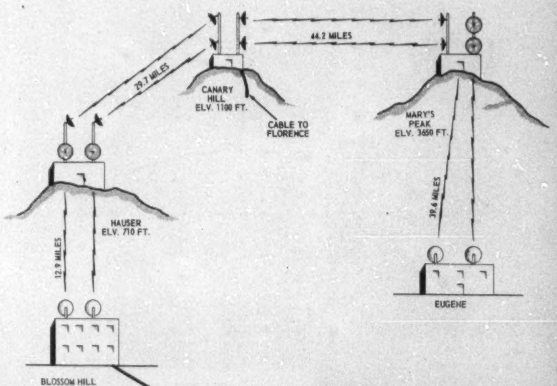


Fig. 4. Eugene–Coos Bay system

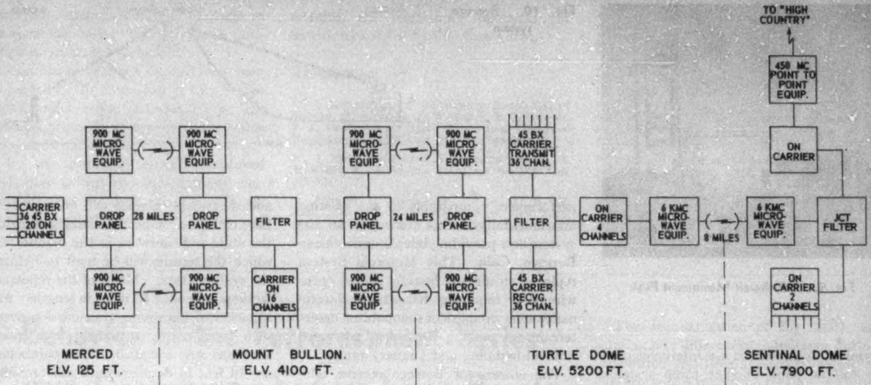

TO "HIGH COUNTRY"
450 MC POINT TO POINT EQUIP.
ON CARRIER
900 MC MICRO-WAVE EQUIP.
45 BX CARRIER TRANSMIT 36 CHAN.
CARRIER 36 45 BX 20 ON CHANNELS
DROP PANEL
28 MILES
FILTER
24 MILES
ON CARRIER 4 CHANNELS
6 KMC MICRO-WAVE EQUIP.
8 MILES
JCT FILTER
CARRIER ON 16 CHANNELS
45 BX CARRIER RECVG. 36 CHAN.
ON CARRIER 2 CHANNELS
MERCED ELV. 125 FT.
MOUNT BULLION. ELV. 4100 FT.
TURTLE DOME ELV. 5200 FT.
SENTINAL DOME ELV. 7900 FT.

Fig. 5. Block layout Merced–high-country route

The former is owned and operated by the West Coast Telephone Company and the latter two by the Pacific Company. Channel-dropping was employed at the Canary Hill repeater to provide circuits from Florence to Eugene in Oregon.

Between Eugene and Canary Hill, the system employs two regular channels and one stand-by, beyond Canary Hill the stand-by and regular channels are provided on a one-for-one basis.

In addition to the through systems, a short Hauser-Coos Bay system was subsequently installed for local service. This system wholly owned by the West Coast Telephone Company is operated without stand-by facilities.

The operation of this through system has demonstrated the ease with which an area formerly remote from the nationwide toll network was interconnected to it by the co-operative venture of the two serving telephone companies.

In May 1956, under agreement with the Department of the Interior, a completely overhauled and modernized telephone system was placed in service in Yosemite National Park; see Fig. 5. The open-wire lead which had served this area for many years was replaced by a 2-hop Lenkurt *72B* radio system operating between Merced, Fig. 6, and Turtle Dome (within the Park) by way of a repeater at Mt. Bullion; see Fig. 7. At Turtle Dome the circuits were extended to a new community dial office (CDO) in the Valley. In addition, a Motorola *MR-10* system was installed between Turtle Dome and Sentinel Dome. At this latter

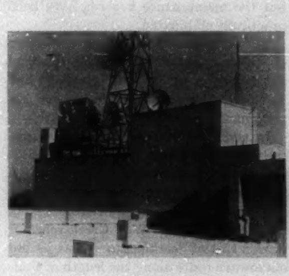


Fig. 6. Merced

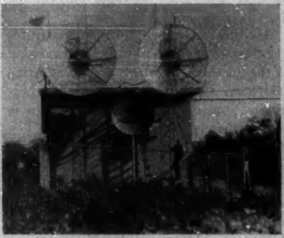


Fig. 7. Mt. Bullion repeater

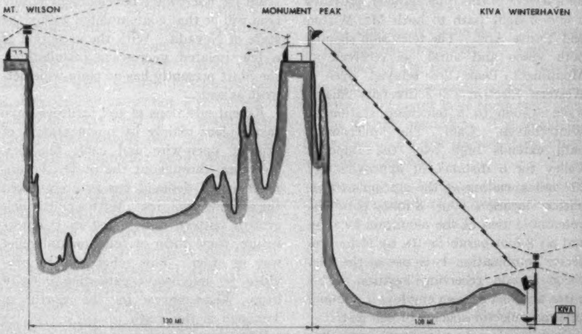

MT. WILSON
MONUMENT PEAK
KIVA WINTERHAVEN
130 MI.
108 MI.
KIVA

Fig. 8. Off-the-air television service

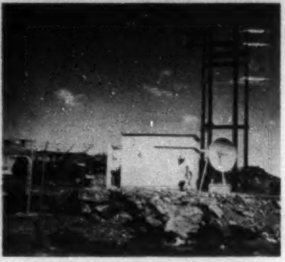


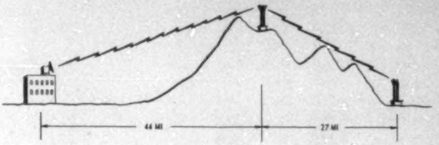
Fig. 9. Off-the-air Monument Peak

terminal, four circuits are interconnected to 450-mc point-to-point radio systems providing telephone service to the "high country." Like so many of the systems installed throughout the mountain regions, this particular installation has brought a service to the area comparable to that previously provided only in the heavily populated valley areas.

Another technique that demonstrates the flexibility of microwave transmission is the method by which a Yuma, Ariz., television station receives network programs; see Fig. 8. Although Yuma is approximately 230 airline miles from Los Angeles, off-the-air television service is provided with only one intermediate repeater. The television transmitters serving the Los Angeles area are located atop Mt. Wilson at an elevation of nearly 6,000 feet. After considerable investigation, a repeater site was chosen atop Mounument Peak, roughly 130 miles south of Mt. Wilson; see Fig. 9. This location in the Laguna Mountains area satisfies the two controlling requirements for a service of this type, namely, provides a line-of-sight path to both Mt. Wilson and Yuma, Ariz. The television signals, both video and aural, as received at Monument Peak are relayed over a Western Electric *TD-2* diversity microwave system to a television station at Winterhaven, Calif. The microwave path extends high over the Imperial Valley for a distance of approximately 100 miles, ending at the station's transmitter location. An 8-foot parabolic antenna is used at the mountain location and an 8-foot parabola–10- by 15-foot reflector combination is in use at the television station receiving location. This latter antenna system employs a nominal antenna-reflector spacing of 400 feet.

Over the years, microwave radio has passed from the status of a child of questionable reliability to an adulthood of performance comparable to any existing instrumentality. One example is the system now providing telephone service to Borrego, Calif. This Motorola System replaced an already heavily loaded open-wire lead requiring hazardous maintenance over 60 miles of uninhabited desert terrain. A portion of this lead traversed a naval-bombing and gunnery range.

Fig. 10. Borrego system



The citizens of Borrego receive telephone service exclusively over a 2-hop microwave system terminating at San Diego; see Fig. 10. Not only has service to the locale been substantially improved, but the maintenance hazards have been practically eliminated.

The foregoing examples are just a few of the continually expanding uses of microwave in the Pacific Telephone and Telegraph Company area to provide improvements in communication facilities.

In addition to these existing systems, the immediate future will see the realization of systems presently in the drawing-board stage. Among these will be a new north-south system between San Bernardino and the Wolf Creek repeater station east of Sacramento. This second microwave route along the length of California will be capable of carrying thousands of telephone circuits or their equivalent. It will provide access to many of the towns in San Joaquin Valley.

Probably the most spectacular program for microwave development in any area will be that contemplated within the state of Nevada. With the exception of a few isolated microwave installations, the State presently has no state-wide network as such.

Recent additions of toll facilities have taken place mainly by augmentation of existing open-wire and cable facilities. However, throughout the State, the unprecedented demand has severely taxed the existing facilities. With a continuing growth pattern anticipated in the near future, some form of fundamental relief was in order. This relief will be provided by microwave extending initially from Wennemucca in the north to Tonapah in the south.

The large cross-section portions of this system will employ Western Electric *TJ* and *TD-2* equipments while the side legs and short-haul sections will make use of Lenkurt units. One interesting feature of this state-wide network is the manner in which the terrain will be used to reduce the system costs. Many of the repeater sections approach 50 miles in length. As a means of improving transmission over such long paths, extremely high gain antennas are specified. The parabolas are 10 feet in diameter and will provide a gain of 41 decibels (11,000) relative to a half-wave dipole. This system will eventually be extended south to Las Vegas where interconnection to the existing Los Angeles–Amarillo route will be possible.

Many of the systems currently in the installation or planning stages will make use of Western Electric *TJ* equipment operating in the 10,700- to 11,700-mc band. The initial application of this equipment is in the Los Angeles area between the Grand Avenue toll office and a similar office in Compton, Calif. Although the path length of 12 miles is short by typical microwave standards, 10-foot parabolic antennas are being used. By providing excess system margins, the channelizing capabilities of the system will be substantially improved. Each microwave channel, so installed, should be capable of transmitting hundreds of voice circuits satisfying a heavy route requirement with short-haul equipments. This system will be used mainly to extend broad-band facilities now terminating at the Grand Avenue toll office to new terminals at the Compton location. In this manner, the circuits are being extended without the requirement for costly back-to-back terminal arrangements.

Conclusions

The period of the last 12 years has seen the expansion of microwave facilities at an almost unprecedented rate. This instrumentality in addition to being economically competitive, provides a flexibility and versatility previously unknown in the communications industry. The broad bandwidths available permit an almost unlimited choice of channelizing techniques. This feature, together with the multiple use of repeater stations, and the ability to modify systems or routes to meet new demands is difficult to assess on an economic basis. However, such flexibility is not being overlooked.

Vast amounts of new capital have been expended in the formation of this Pacific Coast microwave network. For the systems presently in operation the value exceeds 50 million dollars.

By the techniques already employed and those which the future holds, this network will no doubt continue to expand as one of the major methods of meeting the continuing demands for service within the Pacific Coast area.

References

1. A MULTICHANNEL MICROWAVE RADIO RELAY SYSTEM, **H. S. Black, J. W. Beyer, T. J. Grieser, F. A. Polkinghorn.** *AIEE Transactions (Electrical Engineering)*, vol. 65, Dec. 1946, pp. 798–806.

2. A NEW MICROWAVE TELEVISION SYSTEM, **J. F. Wentz, K. D. Smith.** *Ibid.*, vol. 66, 1947, pp. 465–70.

3. MICROWAVE REPEATER RESEARCH, **H. T. Friis.** *Bell System Technical Journal*, New York, N. Y., vol. 27, Apr. 1948, pp. 183–246.

4. FREQUENCY-MODULATION TERMINAL EQUIPMENT FOR THE TRANSCONTINENTAL RELAY SYSTEM, **J. G. Chaffee, J. B. Maggio.** *Electrical Engineering*, vol. 70, Oct. 1951, pp. 880–83.

5. AN UNATTENDED BROAD-BAND MICROWAVE REPEATER FOR THE TD-2 RADIO RELAY SYSTEM, **R. W. Friis, K. D. Smith.** *Ibid.*, Nov., pp. 976–81.

6. THE SEATTLE-VICTORIA RADIO SYSTEM, **R. E. Kistler.** *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 737–42.

7. MICROWAVE FACILITIES WITH BUILT-IN RELIABILITY, **R. G. Kuck.** *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 438–41.

8. PACIFIC COAST TOLL TRANSMISSION NETWORK, **W. L. Carter, R. L Messner.** *AIEE CP 56-782* (not available).

Behavior of the Ferroresonant Series Circuit Containing a Square-Loop Reactor

R. H. DENNARD
ASSOCIATE MEMBER AIEE

Synopsis: An analysis is given for the behavior of the ferroresonant series circuit containing a reactor with 50% NiFe grain-oriented core material. Consideration is given to several different kinds (or modes) of steady-state circuit response at fundamental as well as at subharmonic frequencies for both sinusoidal and square-wave applied voltages. Conditions are derived which are necessary for the stable existence of these modes, and methods are given for the study of the transient processes by which these various steady-state modes are attained.

The dynamic magnetomotive-force requirements of the unsaturated reactor and the differential inductance of the reactor during saturation are represented with suitable approximations. Core test procedures are described which supply this magnetic information.

CIRCUITS combining capacitors and nonlinear inductors exhibit intriguing properties which have been discussed in the literature of the last 50 years. The simple series circuit of Fig. 1 has received particular attention. It is commonly called the ferroresonant series circuit since its bistable behavior was at first considered to be akin to resonance in linear systems. When this circuit is excited with a periodic applied voltage, the steady-state response takes different forms depending on the circuit parameters and on the initial conditions at the time the circuit is closed. This response may be at the same frequency as the applied voltage or at a subharmonic frequency, and the waveforms of current and voltages may or may not be symmetrical with respect to the time axis.

Literature Survey

Until recent years, the ferroresonant circuit contained reactors with core materials of "gradually varying" differential permeability. Analyses for this case have represented the magnetic properties by continuous functions approximating the normal **B-H** curve.[1,2] Many newer materials with greater remanence and sharply defined saturation levels are not suited for that kind of approximation, and representations in the form of straight-line segments have been introduced.

The idealized relationship between flux linkages ψ and current i of Fig. 2(A), with an infinite differential inductance in region (a) and a constant differential inductance in regions (b) and (c), has been used by Travis and Weygandt[3] in their study of subharmonics in the ferroresonant series circuit. The circuit resistance is omitted and the capacitor voltage is assumed to be negligible during saturation.

The approximation of Fig. 2(B), assuming zero differential inductance during saturation, has been applied by Salihi[4] to the study of a mode of operation with the same periodicity as the applied sinusoidal voltage. Brenner[5] uses the same assumption in his study of certain subharmonic modes of operation for various waveforms of applied voltage, modifying Fig. 2(B) by the introduction of a major hysteresis loop with fixed width and of minor loops as shown in Fig. 2(C).

Finally, Evans[6] uses the representation of constant magnetizing current for the unsaturated reactor and of constant permeability μ_0 (of free space) for the core during saturation. With these assumptions he studies a mode of operation having the same periodicity as the applied sinusoidal voltage. A similar approach for the case of zero magnetizing current is found in the unpublished work of R. L. Bright which was presented in a seminar at Carnegie Institute of Technology in 1953. The work of Bright contains the method of studying stability which is incorporated in this paper.

Method of Analysis

This analysis is concerned with values of the circuit parameters for which the inductance of the reactor during saturation and the magnetizing current of the unsaturated reactor have significant roles in the circuit operation. The magnetic properties of the reactor are represented as follows (see Fig. 3):

1. When $|\psi|<\psi_s$ the reactor is unsaturated and the magnitude of the magnetizing cur-

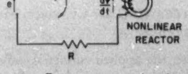


Fig. 1. Ferroresonant series circuit

Paper 58-1215, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 27, 1958; made available for printing August 22, 1958.

R. H. DENNARD is with International Business Machines Corporation, Yorktown Heights, N. Y.

This work is abstracted from a dissertation submitted to the Carnegie Institute of Technology, Pittsburgh, Pa., in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The advice of Prof. L. A. Finzi is gratefully acknowledged. The investigation was performed in part under the sponsorship of the U. S. Office of Naval Research.

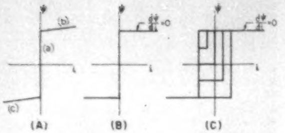


Fig. 2. Linearizations of ($\psi-i$) relationship used by previous authors

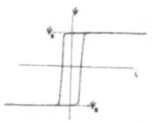


Fig. 3. Typical ($\psi-i$) relationship for square-loop reactor

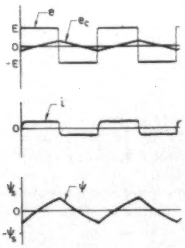


Fig. 4. Waveforms for low-current steady-state mode of operation with reactor A: 1,400 turns of no. 27 American Wire Gauge on 50% NiFe grain-oriented orthonol core of ID = 1 inch, OD = 1.25 inches, height = 0.25 inch, tape thickness = 0.002 inch; E = 17 volts, f = 400 cps, C = 0.1 μf (microfarad), R = 35 ohms

rent is taken to be a constant value I_u which may be evaluated from a core test in terms of the average rate of change of flux linkages, $d\psi/dt$.

2. When $|\psi|>\psi_s$ the reactor is saturated and the differential inductance $d\psi/di$ is an empirical single-valued function of i.

The method of analysis is to solve the circuit equation separately for the unsaturated and saturated conditions, and then to match the solutions at the points of transition.

Analysis for Square-Wave Applied Voltage

Unsaturated (Low-Current) Mode of Operation

When the applied voltage is sufficiently small, the reactor does not reach saturation and the circuit current is the relatively small magnetizing current. Waveforms for this case are shown in Fig. 4. The term iR is negligible in the circuit equation in the unsaturated condition for the values of resistance considered in this paper. Thus,

$$\frac{d\psi}{dt}=e-e_c \qquad (1)$$

using the sign conventions of Fig. 1. The average capacitor voltage during a half-cycle of the applied voltage is approximately zero, so the time integral of $d\psi/dt$ is simply $E/2f$ where E is the magnitude of the applied voltage and f is the frequency. This low-current mode of operation is possible as long as $E/2f$ does not exceed $2\psi_s$. The magnitude of E for which ψ alternates between $-\psi_s$ and $+\psi_s$ is called the "normal" voltage,

$$E_n=4f\psi_s \qquad (2)$$

As E is increased beyond the value E_n, the reactor must become saturated and the circuit operates in a different mode such as the basic high-current mode considered in the following.

Basic High-Current Modes

Fig. 5(A) shows waveforms for a mode of operation in which the circuit current is much larger than in the low-current mode. This is designated as a basic high-current mode because of its relative simplicity.

Consider the operation for an interval in which the reactor is unsaturated as between points *1* and *3* in Fig. 5(A). Equation 1 is valid and shows that $d\psi/dt$ is positive. The magnetizing current is fairly constant so the capacitor voltage increases linearly with time. (Notice, however, the abrupt change of i at point *2* where $d\psi/dt$ changes considerably due to a reversal of the voltage e.) The reactor remains unsaturated until ψ, changing at the rate $d\psi/dt=e-e_c$, reaches ψ_s at point *3*.

While the reactor is saturated ($|\psi|>\psi_s$), as between points *3* and *4*, the circuit behaves as a second-order oscillatory system composed of the saturated inductance, the capacitance C, and the resistance R which includes the winding resistance of the reactor. Actually, the differential equation which governs the circuit behavior is nonlinear since the differential inductance varies as a function of the current. The circuit behavior can be explained qualitatively, however, by assuming that the differential inductance has a constant value during saturation. With this assumption the current is a damped sinusoid which initiates at an almost zero value and has a large positive peak. When the oscillatory current reaches a small negative value at point *4*, the saturated condition is terminated. The saturation interval is short because the current oscillates at a high frequency. During this interval the capacitor voltage changes rapidly and reverses in polarity.

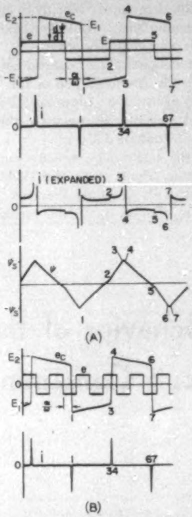


Fig. 5. Waveforms for basic high-current mode

A—At fundamental frequency: E = 17 volts, f = 400 cps, C = 0.1 μf, R = 35 ohms, reactor A

B—For third-order subharmonic: E = 7 volts, f = 400 cps, C = 0.5 μf, R = 20 ohms, reactor A

In the interval between points *4* and *6* the core flux reverses completely. The current is negative in the subsequent saturation interval, and at point *7* the conditions are the same as at point *1*. The operation continues in a cyclic manner.

A similar mode of operation, at a subharmonic frequency, is illustrated in Fig. 5(B) with the labeled points corresponding to the points described in the foregoing. The average magnitude of the reactor voltage, $e-e_c$, is much smaller and the reactor saturates only in every third half-cycle. Other modes of operation may exist in which the reactor saturates only every fifth, seventh, or *m*th half-cycle, where *m* is any odd positive integer and denotes the order of the subharmonic. These are called the basic subharmonic modes. Practical limitations determine a maximum possible value of m as will be shown, but values as large as $m=19$ have been observed experimentally.

A significant quantity which is designated by k and called the "overshoot factor" is here defined as the ratio of the magnitudes of $e-e_c$ just after and before saturation:

$$k=\frac{(e_c-e)_{t=t_u}}{(e-e_c)_{t=t_s}} \tag{3}$$

where t_s is the time at which saturation occurs and t_u is the time at which the reactor again becomes unsaturated. For the basic high-current modes, E_1 and E_2 are the magnitudes of the capacitor voltage just before and after saturation, as shown in Fig. 5, and thus

$$\frac{E_2-E}{E+E_1}=k \tag{4}$$

If the differential inductance during saturation may be approximated by a constant value L_s, then k is equal to $\exp\left(-\pi/\sqrt{(4L_s/R^2C)-1}\right)$. In reality, the differential inductance during saturation is a function of the current, and the value of the overshoot factor k depends on how far the reactor is driven into saturation. k is still a useful notation, and its realistic evaluation is discussed in detail in the next section.

The duration of saturation (which would be $\pi/\sqrt{(1/L_sC)-(R^2/4L_s^2)}$ for a linear saturated inductance L_s) is small in comparison with the unsaturated interval and will be considered zero. The average current I_u during the interval $m/2f$ when the reactor is unsaturated changes the capacitor voltage by the amount $mI_u/2fC$, hence

$$E_1=E_2-\frac{mI_u}{2fC} \tag{5}$$

Equations 4 and 5 give

$$E_1=\frac{1+k}{1-k}E-\frac{1}{1-k}\,\frac{mI_u}{2fC} \tag{6}$$

and

$$E_2=\frac{1+k}{1-k}E-\frac{k}{1-k}\,\frac{mI_u}{2fC} \tag{7}$$

The time integral of $d\psi/dt$ during the unsaturated interval must be $2\psi_s$, and thus

$$\frac{(E_1+E_2)}{2}\,\frac{m}{2f}-E\left(\frac{\pi-\alpha}{2\pi f}\right)+E\left(\frac{\alpha}{2\pi f}\right)=2\psi_s \tag{8}$$

where α is the "firing angle" at which saturation occurs.

These three preceding equations and the definition of equation 2 yield the equation

$$\frac{E}{E_n}=\frac{(1-k)+m^2(1+k)\left(\dfrac{I_u}{4fCE_n}\right)}{(m-1)+(m+1)k+\dfrac{2\alpha}{\pi}(1-k)} \tag{9}$$

relating the firing angle α and the nondimensional factors E/E_n, $I_u/4fCE_n$, and k.

Equation 8 is written with the assumption that the magnitude of the capacitor voltage decreases linearly during the unsaturated intervals. This is approximately true as long as $|e_c|$ is greater than e. It can be shown that this condition is met for any value of $0<\alpha<\pi$ as long as

$$\frac{I_u}{4fCE_n}\leq\frac{2k}{1+k},\ (m=1) \tag{10}$$

or

$$\frac{I_u}{4fCE_n}\leq\frac{k}{m^2(1+k)-k(m-1)-1},\ (m>1) \tag{11}$$

When this condition is met, the mode of operation is also found to be stable as shown in reference 7.

SATURATED INDUCTANCE AND EVALUATION OF k

Part of the flux linkages of the reactor comes from the field in the core material and part from the field in the media surrounding the magnetic core. The linkages of flux in the nonmagnetic media (of $\mu=\mu_0$) are represented as $\psi_0=L_0i$. The flux in the magnetic material contributes flux linkages $\psi_f=N\mathbf{B}_fA_f$ where N is the number of turns of the winding, A_f is the effective cross-sectional area of the toroidal core (as given in the catalogs of core manufacturers), and $\mathbf{B}_f$ is the average flux density normal to that area. The differential inductance is given by

$$\frac{d\psi}{di}=L_0+\frac{N^2A_f}{l_f}\,\frac{d\mathbf{B}_f}{d\mathbf{H}} \tag{12}$$

where l_f is the mean length of the flux path in the magnetic material and $\mathbf{H}$ is the field intensity Ni/l_f.

Fig. 6 shows the relationship during saturation between the differential permeability $d\mathbf{B}_f/d\mathbf{H}$ (relative to μ_0) and $\mathbf{H}$ for three samples of a certain core size. The curves were determined by measuring $d\psi/di$ and calculating $d\mathbf{B}_f/d\mathbf{H}$ from equation 12. The method of measurement was to bias the core into saturation, using a direct current from a high-impedance source connected across the core winding; a small alternating voltage was applied to a series circuit consisting of the same core winding, a known capacitance, and a resistance, and the differential inductance was found by determining the resonant frequency.

The saturated inductance for any reactor with the same core material can be represented by

$$\frac{d\psi}{di}=L_1+L_2g(i/I_2) \tag{13}$$

where $g(i/I_2)$ is a function which is unity for $i=0$ and approaches zero for large values of i. I_2 is a normalizing factor which is arbitrarily defined such that $g(1)=0.1$. Thus, $d\psi/di$ has the value L_1+L_2 for $i=0$ and approaches the value L_1 for large values of i.

A method of evaluating the overshoot factor k is given in Appendix I. The circuit relations are expressed in nondimen-

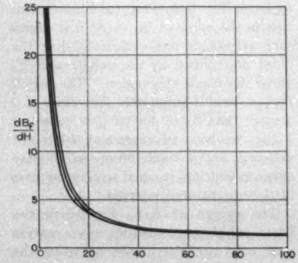
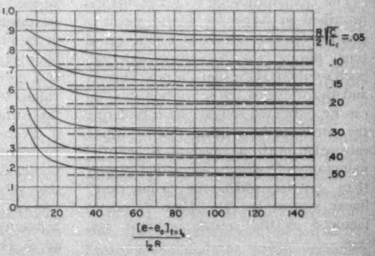


Fig. 6 (left). Relative differential permeability as function of field intensity for core of reactor A and for two other cores of same size

Fig. 7 (right). Overshoot factor k for reactor A; I_2 = 42 milliamperes for this particular reactor



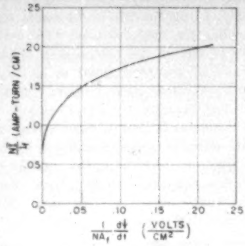


Fig. 8. Results of core test on reactor A for use in evaluation of I_u

sional terms, so as to maintain generality, and the resulting nonlinear differential equation is solved numerically. It is found that k depends on the magnitude of $e-e_c$ at the time of saturation, expressed in the nondimensional term $(e-e_c)_{t=t_s}/I_2R$, as well as on the parameters $(R/2)\sqrt{C/L_1}$ and L_2/L_1. Fig. 7 shows the calculated solution for the case of a particular reactor A having a typical value of L_2/L_1. The dashed lines represent the solution for the linear case in which $d\psi/di=L_1$. For large values of $(e-e_c)_{t=t_s}/I_2R$ it is recognized that **H** has a relatively large peak value during saturation, $d\psi/di$ is nearly L_1 for most of the time, and k is not much larger than for the linear case in which $d\psi/di=L_1$ during the whole saturation interval.

For the basic high-current modes

$$\frac{(e-e_c)_{t=t_s}}{I_2R}=\frac{E+E_1}{I_2R}$$
$$=\frac{2}{1-k}\left(\frac{E_n}{I_2R}\right)\left(\frac{E}{E_n}-\frac{mI_u}{4fCE_n}\right)\quad(14)$$

For a particular situation the value of k can be found by calculating $(e-e_c)_{t=t_s}/I_2R$ for an estimated value of k, locating the corresponding value of k from curves like those in Fig. 7, and repeating the process once or twice if necessary until the value of k found from the curve agrees with the estimated value.

The computation of k by the method described in Appendix I also yields a solution for the current as a function of time during the saturation interval. Such solutions show that the peak current and the width of the saturation interval can be estimated roughly by considering the circuit to be a linear system during saturation with a constant differential inductance of value L_e (somewhat greater than L_1) for which the "equivalent" linear system has the same overshoot factor k as the actual nonlinear system.

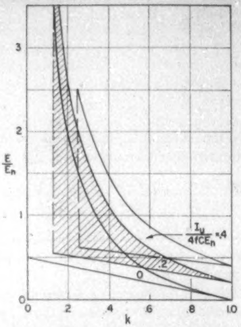


Fig. 9. Regions of parameters for which basic high-current mode is possible

EVALUATION OF AVERAGE MAGNETIZING CURRENT

The value of the average current during the unsaturated condition I_u may be determined from the results of a core test. For a given core it is accepted here that the instantaneous relationship between the magnetomotive force F, the flux ϕ, and $d\phi/dt$ is unique in processes of monotonic flux reversal initiated from pronounced saturation. This relationship can be found conveniently by the procedure of Feth,[8] keeping $d\phi/dt$ constant during flux reversal and measuring the average current I for the interval when the core is unsaturated. The results of such a test for many values of $d\phi/dt$ are given in the curve of Fig. 8; these results are plotted in terms of NI/l_f and $(1/NA_f)(d\psi/dt)$ so that the data can be used to predict the behavior of other cores of the same material and lamination thickness, and of the same ratio of the inside and outside core diameters (ID/OD).

When the ID/OD ratio is reasonably close to unity (about 0.8 or larger), F is fairly constant for a constant $d\phi/dt$, indicating that the influence of the time-varying ϕ is not particularly significant, and results such as shown in Fig. 8 may be used as an approximate instantaneous relationship between i and $d\psi/dt$. Hence in the problem at hand, I_u could be found by integration of the instantaneous values of i corresponding to the instantaneous values of $d\psi/dt$ for the interval in which the reactor is unsaturated. However, if the relationship between i and $d\psi/dt$ is almost linear over the range of variation of $d\psi/dt$, I_u can be approximated by the value of I corresponding to the average value of $d\psi/dt$. For the basic high-current modes the average value of $d\psi/dt$ during the unsaturated condition is $4f\psi_s/m$ and the corresponding values of I give accurate values of I_u as long as $d\psi/dt$ is somewhat greater than zero during the unsaturated interval. The values of I_u decrease for increasing values of m because of the smaller rates of flux change for the subharmonic modes of operation.

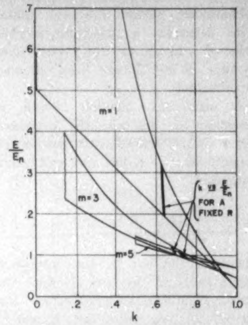


Fig. 10. Ranges of values of E/E_n and k for which various basic modes may occur ($I_u/4fCE_n=0.02$ for $m=1$). Variation of k as function of m and E/E_n is also given for the particular values $(R/2)\sqrt{C/L_1}=0.15$ and $(E_n/2I_2)\sqrt{C/L_1}=11.3$

RESULTS OF ANALYSIS FOR BASIC HIGH-CURRENT MODES

Summing up, it has been seen that the basic high-current modes of operation depend quantitatively on the nondimensional terms E/E_n, $I/_u4fCE_n$, and k. The overshoot factor k actually depends somewhat on the values of the other two terms except in situations where the reactor is driven far into saturation.

The range of parameters over which the basic high-current mode of Fig. 5(A) can exist is limited by the condition $0<\alpha<\pi$. This condition is used with equation 9 to determine the regions shown in Fig. 9. Consider, for example, that $I_u/4fCE_n=0.2$ so that the shaded region in Fig. 9 applies; to obtain operation in the mode of Fig. 5(A) it is necessary (although not sufficient) that the point determined by the values of E/E_n and k lies inside this region. The region for this case is given only for values of k greater than 0.11; for smaller values of k the condition of expression 10 is not satisfied, and a mode of operation may occur in which e_c is equal to e during part of the unsaturated interval.

The ranges of E/E_n for which the basic subharmonic modes of operation may exist are also found by using the condition $0<\alpha<\pi$ and equation 9. The

Vol.
Pt. 1
1958

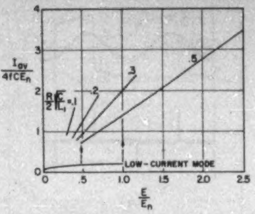


Fig. 11. Average current as function of voltage and resistance for operation in basic high-current mode with reactor A: $I_u/4fCE_n=0.2$ and $(E_n/I_sR)\sqrt{C/L_1}=7.5$

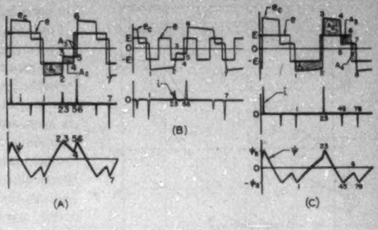


Fig. 12. Waveforms for other modes of operation with square-wave applied voltage

A—E=36 volts, f=250 cps, C=0.8μf, R=10 ohms, 700 turns of reactor A

B—E=14 volts, f=400 cps, C=0.8μf, R=15 ohms, reactor A

C—E=22 volts, f=250 cps, C=0.8μf, R=10 ohms, 700 turns of reactor A

results are illustrated in Fig. 10 for values of k which satisfy conditions 10 and 11. The value of $I_u/4fCE_n$ is taken as 0.020 for $m=1$, and is reduced to 0.016 for $m=3$, and to 0.014 for $m=5$, to account for the smaller values of I_u in the subharmonic modes when the values of the frequency and the capacitance are fixed.

The heavy lines in Fig. 10 show how the overshoot factor k is related to E/E_n and to m for a particular set of circuit values. k is greater for the subharmonics because $e-e_c$ is smaller at the time of saturation. It is seen that for certain values of E/E_n the modes of $m=3$ and $m=5$ are both possible; the low-current mode of Fig. 4 is also possible (since $E/E_n<1$), so the circuit has at least three possible modes of steady-state operation.

The parameter $I_u/4fCE_n$ is particularly important for the subharmonics because of the relatively long unsaturated intervals in which the capacitor voltage decays due to the magnetizing current. Moreover, the average capacitor voltage is much smaller so the amount of decay $mI_u/2fC$ is a much greater percentage of the peak value E_2. Ultimately the rate of the capacitor discharge determines an upper limit on the possible values of m.

The average magnitude of the circuit current I_{av} is given by the relationship

$$\frac{I_{av}}{4fCE_n}=\frac{I_u}{4fCE_n}+\frac{1}{m}\left(\frac{1+k}{1-k}\right)\left(\frac{E}{E_n}-\frac{mI_u}{4fCE_n}\right) \qquad (15)$$

Fig. 11 shows the relationship between current and voltage for various values of resistance when the circuit is operating in the basic high-current mode at fundamental frequency ($m=1$). These curves are for a certain reactor and values of f and C such that $I_u/4fCE_n=0.2$. The curves were prepared by choosing many values of E/E_n and $(R/2)\sqrt{C/L_1}$, evaluating k, and calculating corresponding values of $I_{av}/4fCE_n$ from equation 15. The shaded region of Fig. 9 was used to determine the range of E/E_n for which the mode of operation is possible.

Fig. 11 also shows the average current for the low-current mode of operation. For the case of $(R/2)\sqrt{C/L_1}=0.5$, the circuit behavior is bistable when E/E_n is between the values 0.48 and 1.0. The dependence of the width of the bistable range on $I_u/4fCE_n$ and k is shown by Fig. 9.

The current "jumps up" to the high value when E is raised to just greater than E_n because, once the operation proceeds into the saturated region, the capacitor voltage builds up rapidly and forces the reactor far into saturation repeatedly. The current "jumps down" when the value of E is so low that the capacitor voltage cannot maintain a magnitude sufficiently large to give a resultant average reactor voltage of E_n over a half-cycle.

OTHER HIGH-CURRENT MODES OF OPERATION

A variety of other modes of operation are recognized in the experimentation, and some of these are illustrated in Fig. 12. When the applied voltage is considerably larger than the values for which the basic high-current ($m=1$) mode exists, the reactor becomes saturated twice in each half-cycle as shown in the waveforms of Fig. 12(A). Consider this operation starting first at point *1* where ψ has the value $-\psi_s$. The reactor is unsaturated between points *1* and *2* while ψ increases to $+\psi_s$. The volt-time integral indicated by the area A_1 is thus $2\psi_s$. The reactor is saturated between points *2* and *3*, and the capacitor voltage reverses polarity with respect to the applied voltage e during that time. In the interval *3* to *4*, ψ decreases by an amount equal to the area A_2; at point *4* the applied voltage reverses, changing the polarity of $d\psi/dt$, and ψ increases until the reactor again reaches positive saturation at point *5* which is determined by the condition $A_3=A_2$. In the saturation interval between points *5* and *6* the capacitor voltage again changes rapidly. The waveforms following point *6* are symmetrical to the waveforms starting at point *1*.

A very similar sequence of operation is found in the third-order subharmonic illustrated in Fig. 12(B). In fact, the labeled points correspond to the points described in the foregoing for the mode of Fig. 12(A). Higher order subharmonics also can exist in the same manner.

In the mode of operation illustrated by the unsymmetrical waveforms of Fig. 12(C) the reactor saturates twice in every other half-cycle of the applied voltage. Similar subharmonic modes of operation are also observed. For a certain range of voltages the modes of Figs. 12(A) and (C) are both possible; for somewhat higher voltages only that of Fig. 12(A) can occur and for somewhat lower voltages only that of Fig. 12(C) can occur.

All these modes of operation are investigated in more detail in reference 7, and several other modes of operation are discussed, including some even-order subharmonics. The same general method of analysis given for the basic high-current modes applies also to these many other modes of operation.

TRANSIENT ANALYSIS

The same principles used in the analysis of the steady-state operation of this circuit can be applied also to the study of transients by considering the behavior of the circuit step by step. This method lacks generality but is reasonably quick because the circuit response usually approaches the steady state within a few cycles of the response frequency.

When the circuit is abruptly closed there are three initial quantities which, along with the circuit elements and the applied voltage, determine the ensuing behavior. These quantities are the initial capacitor voltage $e_c]_0$, the initial flux linkages $\psi]_0$, and the phase angle $\Theta]_0$

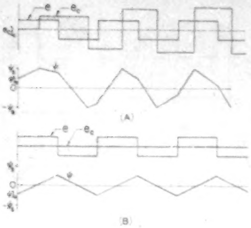


Fig. 13. Typical behavior after closing the circuit for case of E/E$_n$=0.5 and k=0.4

A—$\psi]_0>0$ B—$\psi]_0<0$

of the applied voltage at the time the circuit is closed. As an example, suppose that the circuit is closed with $e_c]_0=0$, $\Theta]_0=0$, and $E/E_n=0.5$. Fig. 13(A) illustrates the processes when $\psi]_0$ is positive under the assumption of negligible magnetizing current. The reactor must become saturated before the end of the first half-cycle in this case (because $\int_0^{\pi/\omega} e\,dt=\psi_s$,) and the time at which this occurs may be readily determined. During the first saturation interval the capacitor voltage rapidly builds up to a value $(1+k)E$ as determined from equation 3 which may be rewritten as

$$(e_c)_{t=t_u}=(1+k)e-k(e_c)_{t=t_s} \tag{16}$$

Subsequently, $e-e_c$ is negative and ψ decreases, reaching $-\psi_s$ at a time such that the shaded volt-time area is $2\psi_s$. The process continues in the same manner with the reactor saturating repetitively and the capacitor voltage builds up, almost reaching a steady-state condition after only 2 cycles of the applied voltage. In this process each firing angle and each value of e_c may be calculated successively. k may be evaluated for each saturation interval, but in this discussion of transients it is assumed that the values of $(e-e_c)_{t=t_s}/I_2R$ are such that k is practically a constant.

For the same conditions but with $\psi]_0$ negative, Fig. 13(B) shows that the reactor is never driven into saturation and the steady-state operation is the low-current mode. Thus it is seen that the initial conditions enter into the determination of the final steady state.

The type of transient shown in the example of Fig. 13(A) is not uncommon; i.e., it occurs frequently also with other initial conditions and circuit parameters. An analysis can be established rather easily in terms of difference equations if the tentative postulation is made that the reactor saturates once and only once in each half-cycle of the applied voltage. (For any particular set of parameters, the results obtained from equation 22 shall indicate whether or not this initial postulation is correct.) E_{c_0} is defined as the magnitude of e_c at the time the reactor is first saturated and α_1 is the corresponding firing angle. E_{c_0} and α_1 can be determined from $e_c]_0$, $\psi]_0$, and $\Theta]_0$ for a particular circuit and applied voltage. $\alpha_{(n)}$ is the firing angle in the nth half-cycle where $n=1$ designates the first half-cycle in which the reactor becomes saturated. It will be assumed that the capacitor voltage decays by the amount $I_u/2fC$ during each interval when the reactor is unsaturated, where I_u is evaluated as for the basic high-current ($m=1$) mode. The average magnitude of the capacitor voltage in the interval between $\alpha_{(n)}$ and $\alpha_{(n+1)}$ is designated $E_{c_{av(n)}}$ and is related to its previous value (for $n>1$) by

$$\left(E_{c_{av(n)}}+\frac{I_u}{4fC}-E\right)=k\left(E_{c_{av(n-1)}}-\frac{I_u}{4fC}+E\right) \tag{17}$$

For $n=1$:

$$E_{c_{av(1)}}=E+k(E-E_{c_0})-\frac{I_u}{4fC} \tag{18}$$

The solution to the difference equation 17 is

$$E_{c_{av(n)}}=E_{c_{ss}}\left[1-k^{(n-1)}\left(1-\frac{E_{c_{av(1)}}}{E_{c_{ss}}}\right)\right] \tag{19}$$

where $E_{c_{ss}}$ is the average magnitude of e_c in the steady-state of the basic high-current mode and is given by

$$E_{c_{ss}}=\frac{1+k}{1-k}\left(E-\frac{I_u}{4fC}\right) \tag{20}$$

The values of $\alpha_{(n)}$ may be determined from the volt-time area relationship:

$$E_{c_{av(n)}}(\pi+\alpha_{(n+1)}-\alpha_{(n)})-E(\pi-\alpha_{(n+1)}-\alpha_{(n)})=2\omega\psi_s \tag{21}$$

The fact that this difference equation is nonlinear discourages attempts at an analytical solution, but for any given case the successive values of $\alpha_{(n)}$ may be calculated from the following recursion formula derived from equation 21:

$$\frac{\alpha_{(n)}}{\pi}=\frac{1-\left(\frac{E_{c_{av(n-1)}}}{E_n}-\frac{E}{E_n}\right)\left(1-\frac{\alpha_{(n-1)}}{\pi}\right)}{\frac{E_{c_{av(n-1)}}}{E_n}+\frac{E}{E_n}} \tag{22}$$

The reactor saturates once and only once in each half-cycle as long as each value of $\alpha_{(n)}/\pi$ is between zero and unity, and this offers a check on the initial postulation.

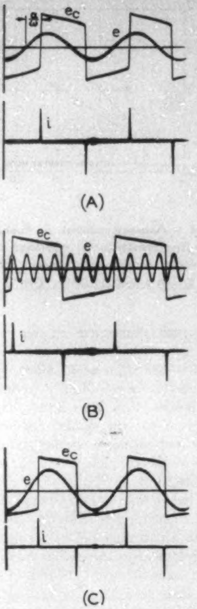


Fig. 14. Basic modes of operation with sinusoidal applied voltage

A—Symmetrical m=1 mode. E=11 volts (rms) f=400 cps, C=0.2 μf, R=15 ohms, 700 turns of reactor A
B—Symmetrical m=7 mode. E=3.2 volts, f=400 cps, C=5 μf, reactor A
C—Unsymmetrical m=1 mode. E=25 volts, f=400 cps, C=0.2 μf, R=15 ohms, 700 turns of reactor A

Equation 19 indicates that the steady-state operation is reached sooner for low values of k. As in the steady-state analysis, the assumption concerning the magnetizing current may be somewhat in error for low values of k but the conclusion is still valid. If $E_{c_{av(n)}}$ is plotted as a discrete function of t, where $t=n\pi/\omega$, it may be shown that the curve passing through these points is identical with the capacitor voltage in an R-C circuit with an applied voltage equal to $E_{c_{ss}}$ and with a value $E_{c_{av(1)}}$ at $t=\pi/\omega$ if the time constant τ of the R-C circuit is

$$\tau=\frac{\pi}{\omega\left[\log_e\left(\frac{1}{k}\right)\right]} \tag{23}$$

This curve approximates an envelope of the magnitude of the capacitor voltage

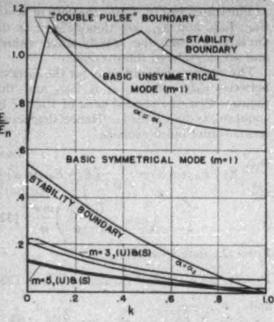


Fig. 15. Ranges of parameters for which basic high-current modes are possible in case of $I_u/4\sqrt{2}fCE_n=0$ for sinusoidal applied voltage

$|e_c|$ for the time following $t=\pi/\omega$, but does not necessarily give any indication of the behavior during the preceding time.

Analysis for Case of Sinusoidal Applied Voltage

The same general method of analysis may be used to study the behavior of the ferroresonant circuit for the case of a sinusoidal applied voltage, $e=E_m\sin\omega t$. Many of the modes of operation are qualitatively similar to the square-wave cases, but the analysis is more involved because the sinusoidal waveform leads to transcendental equations which are tedious to solve.

For this section the symbol E represents the rms value of the applied voltage and E_n is the rms value of the normal sinusoidal voltage ($E_n=4.44f\psi_s$). The low-current mode in which the reactor is unsaturated is possible as long as E is less than approximately E_n.

In the most common high-current mode of operation the reactor becomes saturated once in each half-cycle as shown in the waveforms of Fig. 14(A). During each interval in which the reactor is unsaturated, ψ changes from $-\psi_s$ to $+\psi_s$ or vice versa and the capacitor voltage decreases in magnitude while a small, nearly constant current of average magnitude I_u flows. During each short interval of saturation a relatively large peaked current flows and the capacitor voltage rapidly reverses polarity with respect to the applied voltage $E_m\sin\alpha$ at that time.

In the same manner, subharmonic modes of operation are possible for which the average reactor voltage during the unsaturated condition is so much smaller that the reactor saturates only in every mth half-cycle, where m is odd. An oscillogram for a seventh-order subharmonic of this kind is given in Fig. 14(B).

The waveforms shown in Figs. 14(A) and (B) are symmetrical with respect to the time axis. It is observed experimentally that larger values of the applied voltage may cause the waveforms to become unsymmetrical as illustrated by Fig. 14(C) in which different durations of the unsaturated condition are observed for positive and for negative capacitor voltages.

For the basic symmetrical modes of operation shown in Fig. 14 it can be shown that the firing angle α is related to the nondimensional terms E/E_n, $I_u/4\sqrt{2}fCE_n$, and k by the equation

$$\frac{E}{E_n}=\frac{(1-k)+\frac{m^2\pi}{2}(1+k)\left(\frac{I_u}{4\sqrt{2}fCE_n}\right)}{\frac{m\pi}{2}(1+k)\sin\alpha-(1-k)\cos\alpha} \qquad (24)$$

The overshoot factor k and the average magnetizing current I_u can be evaluated in the manner described for the case of a square-wave applied voltage.

A detailed analysis for the unsymmetrical cases is given in reference 7. The complex relationships cannot be presented in the space available here.

The stability of the symmetrical modes is investigated in Appendix II. It is found that these modes are stable as long as α lies in the range $\alpha_1<\alpha<\alpha_2$, where α_1 and α_2 are the values of α that satisfy equations 42 and 43 respectively. The ranges of applied voltage for which the symmetrical modes exist can be determined by using the stability criteria in conjunction with equation 24. The results are illustrated in Fig. 15 for the case in which $I_u/4\sqrt{2}fCE_n$ is negligibly small, and results for the unsymmetrical mode are also given. (The ranges of applied voltage which are given for the subharmonic modes cover both symmetrical and unsymmetrical operation.) When the capacitor voltage is smaller than the applied voltage during portions of the unsaturated intervals, the reactor is apt to become saturated more often than once each half-cycle; the new mode of operation is called a "double-pulse" mode, and this occurrence limits the ranges of the basic modes as indicated in Fig. 15.

Fig. 16 indicates the effect of the magnetizing-current discharge of the capacitor in the unsaturated intervals for the basic symmetrical modes of operation. The value of $I_u/4\sqrt{2}fCE_n$ is taken as 0.015 for the $m=1$ mode and is reduced to 0.012 for $m=3$ and to 0.011 for $m=5$ to account for the smaller values of I_u associated with the smaller rates of flux change in the subharmonic cases. The regions shown in Fig. 16 are limited to values of k for which the capacitor voltage remains greater in magnitude than the applied voltage during the unsaturated interval; thus the assumption that the magnetizing current is a constant value I_u is a good approximation. It is seen that the fifth-order subharmonic can exist for only a narrow range of applied voltages (corresponding to a narrow range of firing angles for which the mode is stable). For a certain set of circuit values, k is different in the three modes of operation as designated in Fig. 16.

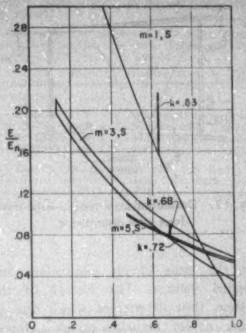


Fig. 16. Ranges of parameters for which basic symmetrical high-current modes are possible for sinusoidal applied voltage

Conclusions

The ferroresonant series circuit containing a square-loop reactor is an intriguing device in view of the large variety of possible modes of operation. The periodic steady-state operation can occur at the same frequency as the supply voltage or at a submultiple of that frequency, odd or even. One or more current pulses can occur within one half-cycle, and alternate half-cycles may or may not be symmetrical.

A detailed analysis has been given for the most common modes of operation, describing these modes quantitatively and determining the ranges of the circuit parameters for which these modes can exist in a stable manner. Overlaps of these ranges show the conditions for polystable operation, i.e., the conditions for which more than one mode can exist

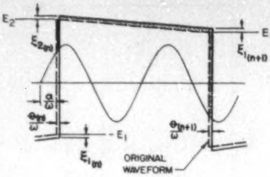


Fig. 17. **Deviations from steady-state waveform after disturbance**

for the same circuit parameters and applied voltage. The history of the system then determines which one of these modes actually occurs. Hence, methods of transient analysis have been given which allow the prediction of the eventual steady state when the capacitor voltage and the flux linkages are known at a given time such as the time at which the circuit is closed.

In the analysis that has been given, the dynamic properties of the reactor during flux reversal are accounted for by defining an average magnetizing current I_u which can be evaluated from the results of a core test. The magnetizing current is particularly significant when the capacitance is small or when the unsaturated interval is long as in many of the subharmonic modes.

The saturated inductance of the reactor is also a significant quantity for which no constant value can be assigned a priori. A method of measuring the saturated inductance is presented, and this leads to a realistic determination of the overshoot factor k in terms of the circuit components and the circuit voltages at the time saturation occurs.

Although the supply frequency used in the experimentation has been in the range of 60 to 1,000 (cycles per second), it appears that the analysis is valid also for much higher frequencies. It is noteworthy that all of the modes of operation found in the experimentation could be explained, or even predicted in advance, using the general concepts that have been established.

Appendix I. Evaluation of the Overshoot Factor k

The circuit equation is

$$(e-e_c)_{t=t_s}=\frac{1}{C}\int_0^t i\,dt+iR+\left(\frac{d\psi}{di}\right)\frac{di}{dt} \tag{25}$$

where time is measured from the beginning of operation in the saturated region ($t_s=0$) and e is considered constant for the duration of the saturated condition. Two nondimensional parameters are defined as follows:

$$x=\frac{t}{RC} \tag{26}$$

$$y=\frac{\frac{1}{C}\int_0^t i\,dt}{(e-e_c)_{t=t_s}} \tag{27}$$

With the use of these definitions and equation 13, equation 25 becomes

$$\frac{d^2y}{dx^2}\left\{1+\frac{L_2}{L_1}g\left[\frac{(e-e_c)_{t=t_s}}{I_2R}\frac{dy}{dx}\right]\right\}\frac{L_1}{R^2C}+\frac{dy}{dx}+y=1 \tag{28}$$

Define

$$\frac{dy}{dx}=z \tag{29}$$

From the previous definition it is seen that $z=iR/(e-e_c)_{t=t_s}$.

Then

$$\frac{dz}{dx}=\frac{4(1-z-y)\left(\frac{R}{2}\sqrt{\frac{C}{L_1}}\right)^2}{1+\frac{L_2}{L_1}g\left[\frac{(e-e_c)_{t=t_s}}{I_2R}z\right]} \tag{30}$$

The differential equation of the circuit has thus been reduced to two simultaneous first-order equations which are subject to the conditions $y]_{x=0}=0$ and $z]_{x=0}=0$ (corresponding to $1/c\int_0^t i\,dt=0$ and $i]_{t=0}=0$). y and z as functions of x can be found by numerical solution of equations 29 and 30 for particular values of $(R/2)\sqrt{C/L_1}$, $(e-e_c)_{t=t_s}/I_2R$, and L_2/L_1, and for a particular function $g(i/I_2)$.

Using equation 27 with the definition of k (equation 3), one finds that

$$k=\frac{(e_c)_{t=0}+(e-e_c)_{t=0}(y)_{t=t_u}-e}{e-(e_c)_{t=0}}=(y)_{t=t_u}-1 \tag{31}$$

From the values of y and z as functions of x, $(y)_{t=t_u}$ may be found as the first peak value of y (at $z=0$).

Appendix II. Stability Criteria for the Symmetrical Basic Modes of Operation; Sinusoidal Applied Voltage

The stability of the modes of operation illustrated in Figs. 14(A) and (B) is studied by considering the effect of a very small disturbance on the system while the circuit is operating in one of the modes under consideration. The firing angle and the capacitor voltage are changed by small amounts. Operation is stable if the system reverts to the same steady state after removal of the disturbance.

Let the nth firing angle after the disturbance be $\alpha_{(n)}=\alpha+\Theta_{(n)}$, and let the magnitudes of the capacitor voltage before and after saturation be $E_1+\xi_{1(n)}$ and $E_2+\xi_{2(n)}$ as illustrated in Fig. 17 for the case of $m=3$. $\Theta_{(n)}$, $\xi_{1(n)}$, and $\xi_{2(n)}$ are thus defined as the deviations from the steady-state values, and are considered infinitesimal.

The average value of $e-e_c$ for the interval between $\alpha_{(n)}$ and $\alpha_{(n+1)}$ is $2\psi_s$, and this requires that the shaded area of Fig. 17 be equal to the dotted area. Hence, neglecting second-order infinitesimals,

$$\frac{\Theta_{(n)}}{\omega}(E_2-E_m\sin\alpha)=\frac{\Theta_{(n+1)}}{\omega}(E_1+E_m\sin\alpha)+\left(\xi_{2(n)}+\frac{I_u}{C}\frac{\Theta_{(n)}}{\omega}\right)\frac{m\pi}{\omega} \tag{32}$$

From the definition of k

$$E_2-E_m\sin\alpha=k(E_1+E_m\sin\alpha) \tag{33}$$

and

$$E_2+\xi_{2(n)}-E_m\sin(\alpha+\Theta_{(n)})=k[E_1+\xi_{1(n)}+E_m\sin(\alpha+\Theta_{(n)})] \tag{34}$$

Since $\Theta_{(n)}$ is infinitesimal,

$$\sin(\alpha+\Theta_{(n)})=\sin\alpha+\Theta_{(n)}\cos\alpha \tag{35}$$

These last three equations yield the relation

$$\xi_{2(n)}=k\xi_{1(n)}+(1+k)E_m\Theta_{(n)}\cos\alpha \tag{36}$$

The capacitor voltage discharges at the rate I_u/C in each unsaturated interval; thus equation 5 is valid for the steady-state mode, and also

$$\xi_{1(n+1)}=\xi_{2(n)}-\frac{I_u}{\omega C}(\Theta_{(n+1)}-\Theta_{(n)}) \tag{37}$$

Substituting the value of $\xi_{1(n)}$ from equation 37 (with $n=n-1$) into equation 36, one finds

$$\xi_{2(n)}=k\xi_{2(n-1)}-k\frac{I_u}{\omega C}(\Theta_{(n+1)}-\Theta_{(n)})+(1+k)E_m\Theta_{(n)}\cos\alpha \tag{38}$$

Equation 32 can be solved for $\xi_{2(n)}$ (and $\xi_{2(n-1)}$), and substitution of these values into equation 38 yields a second-order difference equation in terms of $\Theta_{(n)}$. Equations 5 and 33 may be used to eliminate E_1 and E_2, giving the result

$$\Theta_{(n+1)}+2(A-k)\Theta_{(n)}+k^2\Theta_{(n-1)}=0 \tag{39}$$

where

$$A=\frac{m\pi(1-k^2)\frac{E}{E_n}\cos\alpha+2(1-k)^2\frac{mI_u}{4\sqrt{2}fCE_n}}{4\left(\frac{E}{E_n}\sin\alpha-\frac{mI_u}{4\sqrt{2}fCE_n}\right)} \tag{40}$$

The stability of this mode of operation is determined from the roots of the auxiliary equation

$$M^2+2(A-k)M+k^2=0 \tag{41}$$

Operation is stable as long as the roots M_1 and M_2 are less than unity in magnitude. A detailed examination (using the value of E/E_n from equation 24) shows that operation is stable for $\alpha_1<\alpha<\alpha_2$ where α_1 and α_2 are determined from

$$\tan\alpha_1 = \frac{\dfrac{m\pi}{2}\left(\dfrac{1-k}{1+k}\right)+\dfrac{mI_u}{4\sqrt{2}fCE_n}\times\left[\left(\dfrac{m\pi}{2}\right)^2-1-\left(\dfrac{1-k}{1+k}\right)^2\right]}{1-\dfrac{m\pi}{2}\left(\dfrac{1-k}{1+k}\right)\left(\dfrac{mI_u}{4\sqrt{2}fCE_n}\right)} \tag{42}$$

and

$$|\tan\alpha_2| = \frac{m\pi}{2}\left(\frac{1+k}{1-k}\right),\ \left(\alpha_2>\frac{\pi}{2}\right) \tag{43}$$

References

1. Nonlinear Vibrations (book), J. J. Stoker. Interscience Publishers, Inc., New York, N. Y., 1950.

2. Forced Oscillations in Nonlinear Systems, C. Hayashi. Nippon Publishing Company, Osaki, Japan, 1953.

3. Subharmonics in Circuits Containing Iron-Cored Reactors, Irven Travis, C. N. Weygandt. *AIEE Transactions (Electrical Engineering)*, vol. 57, Aug. 1938, pp. 423–31.

4. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Jalal T. Salihi. *Ibid.*, vol. 75, pt. I, July 1956, pp. 296–307.

5. Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis, Egon Brenner. *Ibid.*, Sept., pp. 450–56.

6. Series Circuit Incorporating a Nonlinear Reactor, R. P. Evans. *AIEE CP 56-683* (available on request).

7. Behavior of the Ferroresonant Series Circuit Containing a Square-Loop Reactor, R. H. Dennard. *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1958.

8. Describing Functions for the Reset of Magnetic Amplifier Cores, G. C. Feth. *Magnetic Amplifiers Technical Report no. 24*, Carnegie Institute of Technology, Aug. 1956.

A New Small Crossbar Telephone System for Private Branch Exchanges

H. H. ABBOTT
MEMBER AIEE

A NEW small crossbar telephone system called the *756A* private branch exchange (PBX) was recently developed for business establishments requiring no more than 60 lines to PBX telephone stations and ten trunks to a telephone central office. Distinguishing features of this new PBX are the small size and attractive appearance of the cabinets for the PBX attendants and the switching equipment. These cabinets have colors and finishes that blend with modern office decor so they can be located satisfactorily adjacent to office furniture.

The floor space occupied by this PBX is smaller than that of earlier designed PBX's of comparable size. The PBX equipment assemblies include plugs and jacks on the interconnecting wires to simplify the work of installing, to shorten the interval from request for service until cutover, and to facilitate the removal and reuse of the equipment in different locations.

The switching equipment is housed in a cabinet as shown in Fig. 1. The cabinet consists of two modules with three slide-in units in each module. It can be furnished in two sizes, one for 40 lines and the other for 60 lines. Growth from the smaller to the larger size can be accomplished by unplugging the right-hand module and substituting another module more fully equipped.

The PBX attendant's equipment usually provided is a console of modern attractive appearance shown in Fig. 2. A cord-type switchboard can be provided for the attendant instead of this console or in addition to it according to the customer's service requirements. A key telephone set, shown in Fig. 3, can be provided remote from the other attendant's equipment, such as at a night watchman's station, to enable the watchman to complete calls during periods of light traffic when the regular attendant is off duty.

Customer Service Features

The service features from the customer's viewpoint are basically the same as those of earlier designed PBX's but are augmented by new arrangements to improve the speed and convenience. An important improvement is the "camp-on" feature. This is provided on incoming calls from the central office to PBX stations when the stations are busy.

An incoming call is indicated by a flashing lamp and an audible signal. To answer this call, the attendant presses a button and lifts her handset to talk. To complete the call, the attendant presses the HOLD button momentarily and then dials the 2-digit number of the desired station. If the station is free, the bell of the station is rung automatically. However, if the station is busy, the camp-on feature comes into operation. This causes the outside calling party to camp on the line to the station and, when the line becomes free, to be automatically connected through to the station immediately without any further attention from the PBX attendant. This assures to the calling party a priority on the busy line and the fastest possible completion of his call. If, while one outside party is camped on a busy station line, a second outside party calls the same line, the second party does not camp on the line and the PBX attendant receives a distinctive signal so that she can report the situation to the calling party.

After a connection has been completed from a trunk to a station, the station user can recall the PBX attendant, for a purpose such as transferring the call to another station, by only a single momentary depression of the station switchhook button. This will cause a flashing signal before the PBX attendant until the attendant answers. Then the attendant

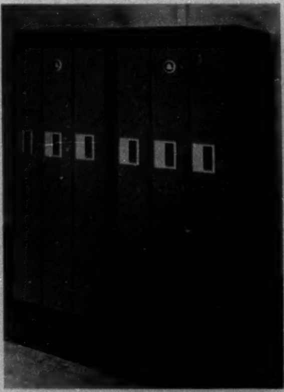


Fig. 1. Front view of switching cabinet

Paper 58-1243, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 13, 1958; made available for printing September 5, 1958.

H. H. Abbott is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

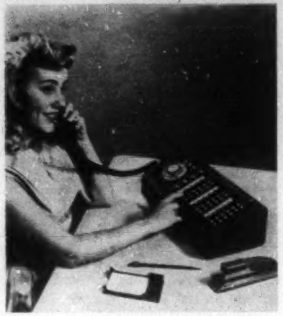


Fig. 2. PBX attendant's console

can proceed with the call as requested by the station user.

When the station user restores his handset without recalling the PBX attendant, the handset depresses the switchhook button continually and, after about one second, the path through the switches is automatically disconnected. The disconnect supervisory signal from the depressed switchhook is delayed in its transmission to the central office via the trunk until the expiration of this one-second interval. This is called the "delayed through supervision" feature.

A call from a PBX station to another PBX station is completed by dialing the 2-digit number of the called station.

Two or more up to a maximum of ten station lines can be arranged in a hunting group so that a call that is directed to one of the lines when it is busy will automatically be connected to one of the other lines in the same group, if one is idle. The numbers of the lines in the hunting group can be either consecutive or nonconsecutive with the same tens digit. The hunting can be optionally two-way or one-way. If it is two-way, a call dialed to any number in a hunting group will hunt to any other number in the same group regardless of whether the other number is higher or lower than the number dialed. If the hunting is one-way, it can be optionally hunted from low to higher numbers, or from high to lower numbers. For one-way hunting, diodes are connected, instead of strap wires, between successive terminals of lines in the hunting group; and the hunting is in the direction of the flow of current through the diodes.

A call from a PBX station via the central office to another station can be completed by dialing the digit 9, waiting for second dial tone from the central office, and then dialing the regular central office number. However, any station line can be restricted from completing such calls. When a line is restricted, a party who dials 9 from the station on this line will hear busy tone instead of the central office dial tone. This restricted service is effective only on outgoing calls dialed directly without assistance from the PBX attendant. The attendant can complete all incoming calls to restricted lines. The calling party can be connected from a restricted line to the central office by dialing 0 instead of 9 and, when the PBX attendant answers, requesting her to complete the connection. Such a connection can be completed while the calling party remains on the line with his handset off-hook.

If the central office is arranged to provide "toll diverting service," any *756A* PBX station line can be restricted from direct dialing of outgoing toll calls without being restricted from local calls. The methods of providing toll diverting service in the various types of central offices, and the limitations of this service in connection with nation-wide direct distance dialing, are beyond the scope of this paper.

When a toll call is completed by the PBX station dialing directly to the toll operator, the connection is held jointly by the station and the operator so that the station can recall the operator and the operator can ring back to the station. Such a ring back will not signal the PBX attendant. However, if the attendant originates a toll call, she will be signaled on a ringback from the toll operator.

Calls originated at PBX stations that are directed to unassigned numbers at the PBX are automatically routed to the PBX attendant.

Night service is now available by either of two methods. One method provides a key telephone set, shown in Fig. 3, for the person who serves as the night attendant. The other method connects certain central office trunks to certain PBX extension lines so that calls on these lines can be originated and received as if they were individual lines to the central office. This latter method is used when no attendant is on duty.

PBX Attendant Service Features

Significant new service features for the PBX attendants are provided in connection with the console shown in Fig. 2 and the key telephone set shown in Fig. 3. These features were designed so that the number of push buttons and lamps would be the minimum for the service requirements, and the size of the attendant's equipment would be the smallest practicable for easy operating.

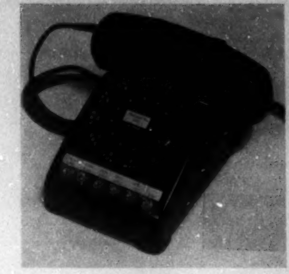


Fig. 3. Key telephone set for PBX attendant

In the attendant's console, each trunk appears on two lamps and a single push-button called a PICK UP button. Two other buttons are provided for use in connection with all of the PICK UP buttons. One is called the RELEASE button. The other is called the HOLD button.

In a fully equipped PBX, 18 PICK UP buttons are provided in the console for the main PBX attendant. Ten are connected to central office trunks and three to attendants trunks. The other five are available for miscellaneous purposes such as to pick up certain PBX lines so that the attendant can answer calls to these lines when the regular users are away. More than one attendant position can be provided for a single PBX to enable the active attendant to be at different locations at different times. An attendant's console may be equipped with only 12 PICK UP buttons instead of 18 where 12 are adequate for the service requirements.

All PICK UP buttons and the RELEASE button are of the 2-position locking type and are mechanically interlocked so that the pressing of any one will release any of the others that were previously operated. These buttons function like those of the familiar 6-button key telephone sets shown in Fig. 3. The RELEASE button is pressed by the attendant when she wishes to disconnect from a trunk without connecting to another trunk. If she does wish to connect to another trunk, she merely depresses the PICK UP button of this other trunk.

When the HOLD button is operated momentarily, it produces two results simultaneously: It places a HOLD condition on the central office trunk, and it connects the attendant to a register circuit with

VoL.
Pt. 1
195

dial tone so that she can dial the 2-digit number of the PBX station line desired by the outside calling party. The HOLD button is not mechanically interlocked with the PICK UP buttons.

The HOLD button is also used to release a connection through the PBX switches on infrequent occasions such as when the attendant makes a mistake in dialing and must release the connection and start over; when the called line is busy and the calling party does not wait, and so it is necessary to release the camp-on condition; and when a call from a trunk to a PBX line is transferred to another PBX line, and it is necessary to release the connection to the first line. This releasing function of the HOLD button is accomplished by successive operations. The first operation of the HOLD button causes a connection, the second operation causes a disconnect, the third another connection, the fourth a disconnect, etc. The attendant knows whether the connect or disconnect conditions exist by the visual signals from the lamps and the tone signal heard in her telephone set. The use of this single button to hold and also release is an important element of the design to enable the provision of small, compact, and simple pushbuttons and lamps for the PBX attendant.

The two lamps associated with each central office trunk provide distinctive meaningful signals. One of these lamps indicates the signals related to the central office side of the connection between the central office trunk and the PBX station line. The other lamp indicates the signals related to the PBX station side of the connection. Thus, these two lamps are comparable with the front and back supervisory lamps of a cord circuit in a cord-type switchboard.

Five different signals can be provided by each of these lamps: (1) dark, (2) steadily lighted, (3) flashing once every 2 seconds, (4) flashing once every one second, and (5) flashing twice a second. The slowest of these three flashes is called a wink signal because the flash consists of a very short dark period and a relatively long lighted period. A steadily dark lamp indicates that the circuit is idle. A steadily lighted lamp indicates that the circuit is busy. A flashing lamp indicates a call in progress. The speed of the flash is in direct proportion to the urgency of the action required of the PBX attendant. A flash at the highest speed (twice per second) indicates that the attendant should take action promptly. This occurs on an unanswered incoming call from the central office, on an unanswered recall after the completion of a call to a PBX station, and on a call to a busy line when another call is camped on the line. A flash at the intermediate speed (once per second) indicates that the called PBX station line is busy and that the trunk is in the camp-on condition waiting for the line to become free. The slowest flash (a wink once every 2 seconds) indicates either one of two different conditions: When this wink is on the trunk lamp, it indicates a HOLD condition on the trunk; when the wink is on the station lamp, it indicates that a connection has been completed through the switches to the station line and that the station bell is ringing awaiting the answer of the called party.

A "split" connection which enables the attendant to announce a call to a PBX station without the outside party listening in is automatically established when the attendent first completes the call. The "split" is automatically eliminated when the attendant disconnects. If she subsequently returns, the elimination of the split is continued, and she is then bridged on the connection; except that, if desired by the customer, the circuits can be wired to produce the split on all occasions while the attendant is connected so as to assure that the attendant cannot listen on conversations between PBX stations and outside parties.

Where the 6-button key telephone set, shown in Fig. 3, is provided for a PBX attendant, the method of operation is substantially the same as that described for the attendant's console of Fig. 2 except that only one lamp instead of two lamps is provided with each PICK UP button, and a switchhook button is provided instead of a RELEASE button. With only a single lamp, the attendant does not know whether the condition producing a flashing lamp is on the trunk side or PBX station side of the connection. However, this absence of a distinctive

Fig. 4 (left). Close view of switching apparatus in 756A PBX

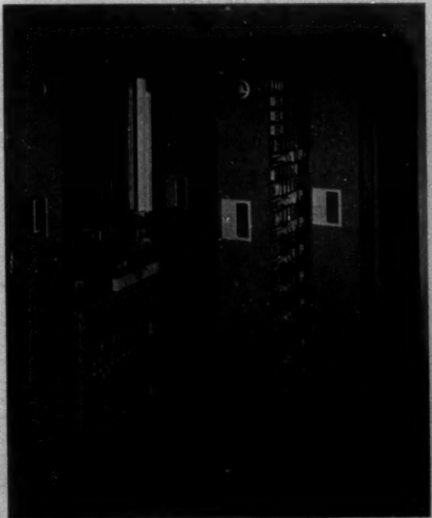


Fig. 5 (right). Front view of switching cabinet with slides 2 and 5 pulled out

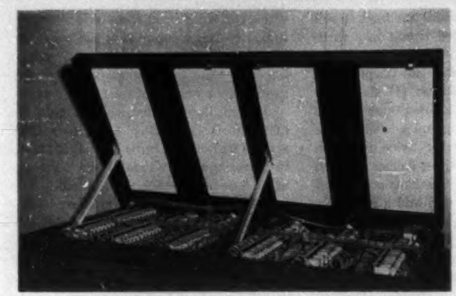


Fig. 6 (left). View of cables in top of switching cabinet with covers open

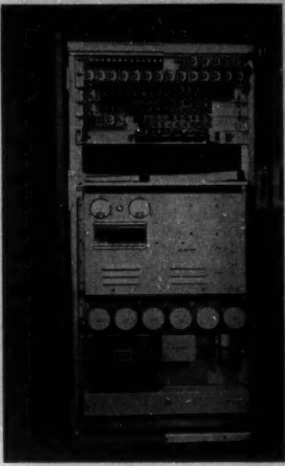


Fig. 7 (right). Batteryless power plant on slide 1 of switching cabinet

indication of the source of the call is not serious where the amount of traffic is small.

Switching Apparatus

Most of the switching apparatus consists of relays of the wire spring type[1] and crossbar switches. A close view of this apparatus is shown in Fig. 4. These relays and switches are substantially the same as those in the latest types of crossbar central offices. The contacts on this apparatus that open and close the electric circuits are of the twin type that minimize troubles from dirt.

Each of the crossbar switches in this PBX has ten hold magnets, ten select magnets, six pairs of make contacts at most of the crosspoints, and two pairs of break contacts on each of the hold magnets.

Some of the wire spring relays are of the new "dual armature" type that provides two independent relays, each with as many as five sets of springs, in the same space required for a single relay with 12 sets of springs.

Equipment and Wiring Features

The external appearance of the cabinet housing the switching equipment is shown in Fig. 1. Fig. 5 is a view of the equipment cabinet with two of the sliding racks pulled out to the position providing access to the switching apparatus for maintenance purposes. The crossbar switches and wire spring relays are similar to those in no. *5* crossbar central offices.[2] Most of the wires to the terminals of this apparatus are attached by solderless wrapped connections like those also used in the latest types of central offices.

The structural framework holding these apparatus slides was carefully designed and tested to produce adequate strength with a minimum amount of material. Tests indicated that this framework will withstand a force of 2 tons applied from one corner to the furthest opposite corner. The apparatus racks are of aluminum to keep the total weight to a minimum. These racks are mounted on ball bearing slides so that they can be readily pulled out and pushed in. An interlocking mechanism among the three slides on each module prevents more than one of the three slides from being pulled out at any one time. This is a safety feature to avoid tipping of the cabinet that might otherwise occur if two slides were pulled out simultaneously. It also eliminates the necessity for fastening the cabinet to the floor.

The slides and modules are interconnected by a cable terminated in multicontact plugs located in the top of the cabinet as shown in Fig. 6. Access to these cables is provided by lifting the covers. The wires to the outside are also connected through plugs and jacks in the top of the cabinet.

The wires from the movable apparatus racks to the fixed plugs in the top of the cabinet are in a looped cable form held together by a canvas cover with a zipper fastener.

Power Supply

The power supply for this PBX is basically 48 volts d-c, capable of supplying 7.5 amperes continuously with momentary peaks up to 12 amperes. Most installations are furnished without batteries. However, electrolytic capacitors of 30,000 microfarads are provided to store power for a fraction of a second so that connections through operated relays and switches will not release during momentary failures or large voltage dips in the power supply. The batteryless type of power plant is located on slide 1 of the switching cabinet as shown in Fig. 7. In locations such as police and fire departments and hospitals, where continuous service is essential even during power failures of more than a fraction of a second, a battery can be added to the power plant and located in an adjacent cabinet.

In the batteryless type of power plant the 20-cycle voltages for ringing bells and the tones for busy and audible ringing signals are produced by saturable-reactor circuits driven by 60 cycles a-c. These are interrupted by cams driven by a 60-cycle a-c motor.

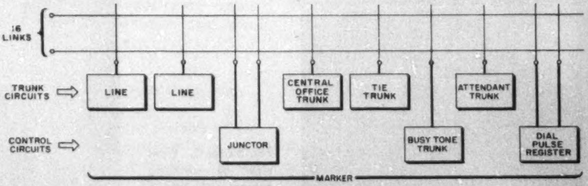


Fig. 8. Switching plan of 756A PBX

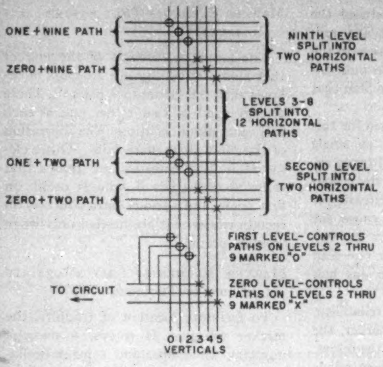

ONE + NINE PATH
ZERO + NINE PATH
NINTH LEVEL SPLIT INTO TWO HORIZONTAL PATHS
LEVELS 3-8 SPLIT INTO 12 HORIZONTAL PATHS
ONE + TWO PATH
ZERO + TWO PATH
SECOND LEVEL SPLIT INTO TWO HORIZONTAL PATHS
FIRST LEVEL-CONTROLS PATHS ON LEVELS 2 THRU 9 MARKED "O"
TO CIRCUIT
ZERO LEVEL-CONTROLS PATHS ON LEVELS 2 THRU 9 MARKED "X"
0 1 2 3 4 5
VERTICALS

Fig. 9. Method of deriving 16 links from ten horizontal paths through the crossbar switches

In the battery type of power plant, transistor oscillator circuits are provided for generating the alternating voltages that produce ringing and tones and drive the motor of the interrupter. These transistor circuits operate entirely from the direct current supplied from the battery so that normal service will not be affected by a prolonged failure in the source of 60-cycle power. To safeguard service in case of failure of a transistor, two transistor circuits are provided for 60 cycles to drive the motor of the interrupter, and two other transistor circuits are provided for the 20-cycle ringing. One circuit of each of these two pairs is normally in service; but in case of failure, the other is automatically transferred into service, and an alarm is indicated. Only a single transistor circuit is provided for dial tone and busy tone because failure in these tones is not as serious as a failure in ringing.

Switching Plan

The switching plan uses nine crossbar switches in a fully equipped PBX, arranged as indicated in Fig. 8. The lines, trunks, junctors, and registers are connected to the verticals. The links are connected to the horizontals. Each of the nine switches has ten verticals. There is a total of 90 verticals accessible to all of the links on the horizontals.

These switches are wired so that 16 links are derived from the ten horizontals. This is accomplished by dividing the horizontals into two groups with eight in the upper group and two in the lower group, as shown in Fig. 9. On each of the eight horizontals in the upper group, there are two links. This provides a total of 16 links. One of these 16 links is selected by choosing one of the eight upper horizontals in combination with one of the two lower horizontals.

Each of the links has three wires. These are connected to a vertical by the closure of contacts at the point where the horizontal crosses the vertical. As there are two links on every horizontal, and three wires per link, there are six wires per horizontal and six contacts per crosspoint on each of the eight upper levels. Only three of the six wires are connected to crosspoint contacts on the first level, and the other three of the six wires are connected to crosspoint contacts on the second level. The connection from a circuit on a vertical to a link on a horizontal is completed through a crosspoint closed on the first or second level in tandem with another crosspoint on the same vertical closed through a crosspoint on one of the upper eight levels.

For each of the ten horizontals there is one select magnet on each of the nine switches. For each of the 90 verticals, there is one hold magnet. These magnets control the connections through the crosspoint contacts between the vertical and horizontals. To complete one of these connections, such as from a line to a link, the marker performs the following four steps in succession:

1. Connects to all the select magnets on the switch containing the vertical to be operated.
2. Operates the two select magnets that identify the link, one on one of the lower two horizontals, the other on one of the upper eight horizontals.
3. Operates the hold magnet.
4. Disconnects from all the magnets.

After the marker disconnects, the select magnets release, but the hold magnet is held operated via the crosspoints and the link.

Basic Circuit Operations

When a customer at a PBX station lifts the handset to originate a call, a line relay operates and summons the marker, as illustrated in Fig. 10. This causes the connections to be completed from the station line via an idle link to a dial pulse register from which dial tone is supplied to the station. When the customer dials the desired number, the dial pulse register circuit receives the dial pulses, translates them into the routing desired, and directs the marker to complete the connection through the crossbar switches to the desired point. Then the register circuit is released.

On a connection from one PBX station line to another station on the same PBX, the connection is completed via two links and one junctor as marked by the X's in Fig. 10. Six junctors are provided. They supply talking, supervisory, and

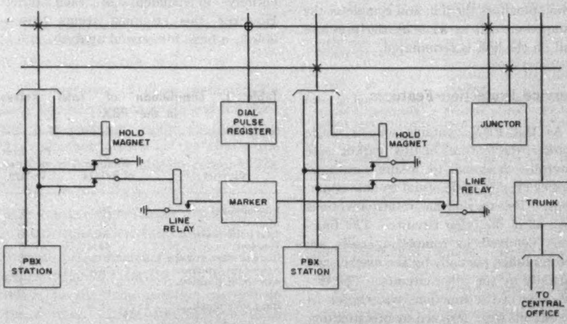


Fig. 10. Basic circuit operation of 756A PBX

ringing currents to the stations. Each junctor is connected to two verticals. One vertical serves the calling line and the other the called line.

On a connection between a central office trunk and a station line, only one link is used. When an incoming call occurs on a trunk, the call is answered by the attendant connecting directly to the trunk through contacts of a relay operated from the pick-up key of the trunk. The attendant completes the call by operating the HOLD button momentarily to cause a connection from her telephone set via the trunk and a link to a dial pulse register. After the attendant has dialed the number of the called station line, the connection is completed via the same link to the line and the dial pulse register is released. Then the ringing current to the station bell is supplied from the trunk circuit.

If the line is found busy when it is called from a trunk, the camp-on condition is produced in the trunk circuit as a result of certain tests made by the marker. The circuits for the camp-on condition make use of three different voltages that may be found on the sleeve lead of each link under different conditions. The first of these is −48 volts that is placed on every idle link by the marker during link testing. The second is −5 volts that is normally on every busy link. The third is 0 volts that is derived from direct ground connected by the marker to the sleeve of a line circuit after the line has been found busy. This direct ground on the line sleeve appears also on the link sleeve to which it is connected. The link test circuit in the marker searches for this direct ground to locate the link associated with the busy line from among the other links that have either −5 volts or −48 volts. The marker then connects the calling trunk to the proper link and establishes the camp-on condition in the trunk. A diode sensing circuit in the trunk monitors the link and completes the camped-on call as soon as the previous call on the link is terminated.

Service Protection Features

As this PBX contains only a single marker, the features of this marker were carefully designed to assure continual service in spite of the usual trouble conditions. This marker has interconnections with all of the other circuits. The functions involved in completing calls are accomplished partially by the marker and partially by the other circuits. The distribution of the functions was chosen as the circuits were designed by investigating various possible plans and then choosing the particular plan which produced the optimum combination of cost and reliability. The plan chosen was not the cheapest nor the most reliable but was judged to be the least expensive plan that would yield reliable operation.

The criterion of reliability used for this common control design was: "A single trouble of any one of the usual types experienced with relays, such as dirty contacts, crosses, grounds, open circuits, or faulty adjustments, shall not cause the failure of service on more than ten station lines, nor on more than one half of all links, nor on more than one half of all trunks in any one group."

To meet this criterion of reliability, the number of relays in the marker, the number of contacts on many of the relays, and the number of wires between relays is greater than that which would be required to provide the same traffic operating features if reliable service under trouble conditions were disregarded.

The extra cost of making the *756A* PBX reliable, above the minimum cost for merely operating under the most favorable conditions, is about 10 per cent of the total cost.

The relays in the *756A* PBX constitute the major portion of its cost. There are 731 relays in a basic *756A* PBX for 60 lines. These relays are distributed among the various circuits as shown in Table I.

FEATURES OF CENTRAL OFFICE TRUNK

Many of the service features derived from the central office trunks are produced by relays individual to the trunks. Other features are produced by relays in the marker and position circuits that are used on a common control basis for different calls on all of the trunks.

The features produced by the relays in these trunks are illustrated by Table II. These eight groups of functions are not entirely independent of each other. However, this grouping seems reasonable as a basis for general analysis.

MARKER ELEMENTS THAT ASSURE RELIABILITY

The principal elements of the marker that assure reliability are 75 pairs of relays called the "assurance pairs." These are arranged so that either one of each pair can alone continue the operation of the PBX if its mate fails. Other elements in this category are relays that cause second trials if failures occur on first trials, and various extra contacts on certain relays that are needed only when troubles occur.

MARKER ELEMENTS THAT FACILITATE LOCATION OF TROUBLES

To facilitate location of troubles, the marker contains 24 relays, 4 message registers, 12 lamps, and some miscellaneous contacts on certain other relays. These contacts interconnect with these relays and lamps so that an alarm is indicated if a trouble occurs. The different lamps help to locate the source of the trouble.

MARKER RELAY PAIRS THAT ASSURE RELIABILITY

If one relay of the 75 assurance pairs should fail to operate when its mate operates, an alarm is produced while service from the customer's viewpoint is continued without interruption. This method of duplicating relays serves in the *756A* PBX as the counterpart of two or more markers in the larger common control switching systems. In these larger systems, when a trouble in a relay blocks the progress of the marker, the block continues for a short time interval and then the other marker is automatically brought into service to complete the connection. In the *756A* PBX when a trouble occurs in one of these 75 pairs of relays, the marker continues to complete the call without delay and only an alarm condition is produced. Theoretically, it would be possible for several different troubles to occur simultaneously at

Table I. Distribution of Total Relays in the PBX

Circuits	Number of Relays	Per Cent of Total Relays
Line circuits	70	10
Marker	329	45
Dial pulse registers	58	8
Junctors	24	3
Central office trunks	210	28
Attendant trunks	24	3
Attendant position	4	1
Alarms	11	2
Busy tone trunk	1	–
Total	731	100

Table II. Functions of Relays in Central Office Trunk Circuit

Function	Number of Relays
Ringing from central office	2
Control signals to and from central office including delayed through supervision	5
Connector for leads to attendant's position	2
Holding central office, inward completing, and releasing	5
Machine ringing	2
Camp-on	3
Flashing recall	1
Toll diverting	1
Total	21

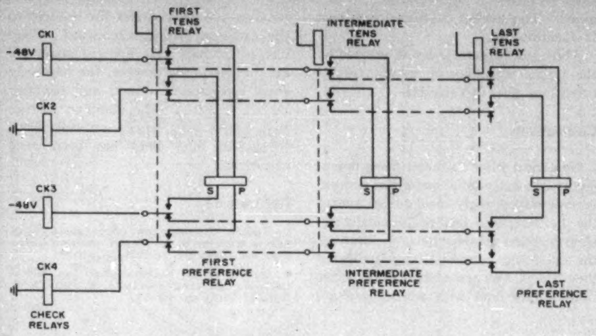


Fig. 11. Protective features of the tens and preference circuits of the marker

several points without interrupting service from the customer's viewpoint. However, the basic design objective has been to assure service only if one trouble occurs at a time.

The reaction on customers' service that could result if the functions of these 75 pairs of relays were performed instead by only 75 individual relays can be estimated approximately by considering the troubles experienced on similar relays in crossbar central offices. These troubles average about 0.5 per 100 relays per year. The actual number of troubles is influenced by the number of operations and the environment. However, this figure seems reasonable for the general analysis here.

The omission of these 75 relays from the marker would have left the marker vulnerable to a service failure any time a trouble occurred on any one of the remaining 75 relays. Thus, on the basis of an expectancy of 0.5 troubles per 100 relays per year, there would be an expectancy of a service failure of $0.5 \times 75 \div 100 = 0.375$ failures per PBX per year, or a failure of every PBX about once every 3 years. Each of these failures could destroy all service through the PBX for a considerable time until a repairman reached the PBX on the customer's premises and found the trouble. Such a frequent occurrence of complete failures would be substantially below the standard of reasonably good service from the customer's viewpoint. Therefore, these extra relays are provided in the standard product.

Extra Marker Contacts That Assure Release of Certain Relays

Other elements of the marker that assure reliability are a considerable number of extra contacts on certain relays that are arranged so that the release of certain other relays is caused by opening the circuits to both sides of a winding, one side going to ground, the other side to battery. For simple control disregarding trouble, it would be necessary to open the circuit on only one side of the winding. This principle of duplicating contacts assures the release of the relays even though one of the contacts be accidentally short-circuited. This is important in certain situations such as where the relays connect leads from individual trunks to the common control circuits and, if such relays should be accidentally operated, they would cause serious interference with normal service on many trunks.

Extra Marker Contacts That Assure Operation of Chain Circuits

Other elements of the marker which assure reliability are extra contacts on relays in the three principal chain circuits:

1. The tens selection chain.
2. The units selection and line and trunk hunting chain.
3. The link selection chain.

The first two of these chain circuits have two selecting channels and the third has four entrance points, so that, if a dirty contact occurs in the chain, the marker will still be able to complete the call.

Marker Lockout Relays That Assure Rotational Serving of Calls

Some other elements of the marker to assure reliability are the TA- and TRA-relays in the tens selection portion of the marker. There are eight of these relays for 60 lines, and six for 40 lines. They assure rotational selection of all calls waiting to be served by the marker even though a trouble in any one line or trunk circuit produces a condition that continually calls for the marker and tends to monopolize its use. These TA- and TRA-relays are important to reliability because if they were omitted, a failure on any one of the relatively larger number of line and trunk circuits could block service to nearly all other lines or trunks in the PBX.

Tens and Preference Circuits of the Marker

The tens and preference circuits in the marker control the sequence in which calls are served when several are simultaneously directed to the marker. The calls are handled via the tens group identifying circuit and the tens group preference chain that are illustrated in Fig. 11 in simplified form. These circuits have significant safety features because, if they were otherwise, a trouble in them could disable the PBX.

When a call is waiting to be served by the marker, one of the tens relays is operated.

To guard against the more common types of trouble such as open circuits or leads crossed to ground or battery, two separate windings are furnished on each preference relay. Both of these windings are normally open at contacts of the corresponding tens relay. The other contacts on the tens relays are wired in a series

Table III. Typical Relay Troubles and Remedial Action

Kind of Trouble	Circuit Action
Primary winding of a preference relay becomes open	Check relays *CK1* and *CK4* fail to operate, producing alarm. Preference relay operates on its other winding
Primary winding of a preference relay is falsely grounded	*CK4* will be short-circuited and will not operate. This produces an alarm, but the preference relay operates on its other winding
False ground on an upper normally closed contact of a tens relay	*CK1* operates at once to produce alarm. Preference relay will operate on other winding
False ground on a lower contact of a tens relay	*CK2* does not operate when *CK4* operates and this produces alarm. Preference relay operates on both windings
Tens relay fails to operate	Lines associated with that tens group cannot make outgoing calls. Other groups are not affected
First or intermediate tens relay operates falsely	A timing relay (not shown) operates after a short interval to force the release of this tens relay. Other tens relays can then operate their preference relays

chain arrangement so that only one preference relay will operate at a time even though more than one tens relay operates. The check relays *CK1* and *CK4* operate in series with one winding of the preference relay involved on each call. The check relays *CK2* and *CK3* operate in series with the other winding of the same preference relay. All four check relays operate in series with a preference relay each time a call is made. That is an indication that all is well. If one or more check relays operate and at the same time one or more other check relays of the four do not operate, either before a call starts or during progress of a call, an alarm is given. The marker, however, continues to function.

These safety features are illustrated by the typical examples of specific trouble conditions given in Table III.

Conclusions

New *756A* PBX's are now being manufactured in substantial quantities and will become increasingly prominent among the dial telephone facilities at moderate-sized business establishments. Many of the operating personnel have indicated they believe this new system to be better than former ones with which they are familiar and they prefer the console to the former cord-type switchboard. They like particularly the modern, attractive appearance of the console, the relatively small equipment cabinet, and the new service features. The camp-on feature that comes into service on calls from trunks to busy lines has been very attractive.

References

1. DESIGN FEATURES OF BELL SYSTEM WIRE SPRING RELAYS, H. M. Knapp. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 482–86.

2. RECENT NEW FEATURES FOR THE NO. 5 CROSSBAR SWITCHING SYSTEM, J. W. Dehn, R. E. Hersey. *Ibid.*, pp. 457–65.

Assessment of Effects of Delay Distortion in Data Systems

A. D. FOWLER
NONMEMBER AIEE

R. A. GIBBY
NONMEMBER AIEE

AN ESSENTIAL objective of a pulse transmission system is the preservation of some aspect of the pulse shape at the receiver. The phase characteristics of a system, which can generally be neglected in voice transmission, become important in pulse systems, since the shape of a pulse is influenced greatly by the phase of the energy components of the pulse spectrum. A quantity frequently used in describing this influence is the "group delay" or "envelope delay." It is a measure of the relative phase shift experienced by neighboring components of the frequency spectrum and is equal to the slope of the tangent to the phase curve at any given frequency. If the phase shift changes linearly with frequency, the envelope delay is constant across the band, and there is no delay distortion. When the envelope delay varies with frequency, however, the shape of a pulse is degraded or distorted in transmission. This variational component of envelope delay will be called the delay distortion.

When the delay distortion of a transmission system becomes large beyond a certain maximum value in the energy band of a pulse signal, the performance of the system becomes unsatisfactory. For smaller values of distortion, satisfactory performance may be possible but the signal is transmitted with a reduced margin against failure due to noise.

In practice many different delay distortion characteristics are encountered in transmission lines, depending on the type and length of line, number of repeaters, etc. It would be an almost impossible task to take all of these characteristics into account unless some organized approach were used. It would be helpful if the degradation produced by delay distortion could be considered in the same way as degradation due to noise. This approach is complicated, however, by the fact that delay distortion cannot be represented as a single number (such as rms power of a given noise spectrum), but varies across a band of frequencies.

It can be shown that for voice-frequency transmission circuits, many delay distortion curves have the common property of approximate parabolic shape about a vertical axis. Since a parabola is defined by three points, it is possible to represent the delay distortion approximately by the values of delay at three frequencies. These three frequencies are also useful in defining two parameters which can be used to designate a given parabolic delay characteristic. The manner of defining these two quantities is shown in Fig. 1. One parameter is the slope, M, of a straight line intersecting the parabola at the lower frequency f_1 and the upper frequency f_3. The other is the sag, S, which is the distance at mid-frequency, f_2, between the parabolic curve and the straight line defined in the

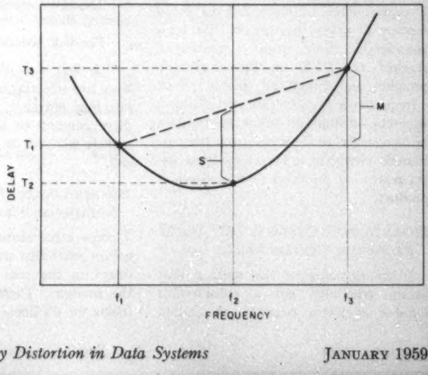
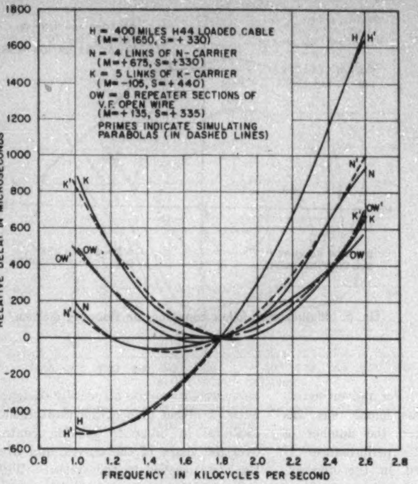


Fig. 1. Slope and sag designations

Paper 58-1240, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 25, 1958; made available for printing September 8, 1958.

A. D. FOWLER and R. A. GIBBY are with Bell Telephone Laboratories, Inc., Murray Hill, N. J.



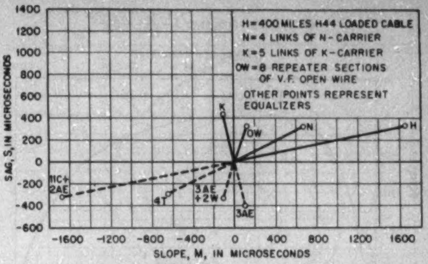


Fig. 2 (left). Parabolic approximation for delay-frequency characteristics message facilities

Fig. 3 (above). Slope-sag plot of message facilities and 26 type equalizers

reference to slope. Any parabola having symmetry about a vertical axis can be designated by the two parameters M and S, as illustrated in Fig. 1.

Using the three frequencies $f_1=1{,}200$, $f_2=1{,}800$, and $f_3=2{,}400$ cps (cycles per second), the delay distortion curves for many voice-band facilities can be represented by parabolas with discrepancies no longer than about 50 μsec (microseconds) over a range from 1,000 to 2,600 cps. This accuracy is within the capabilities of most existing delay-measuring equipment covering this frequency range. The parabolic approximations of delay for four message circuits are compared with the actual delay distortion characteristics in Fig. 2.

The slope and sag of a parabola do not depend on the absolute delay, and will not be altered if an arbitrary value of delay is used as reference. If the delay at frequency f_2 is chosen as reference, the quantities M and S can then be defined in terms of the delay values as

$$M=T_3-T_1$$

$$S=\tfrac{1}{2}(T_3+T_1)$$

where T_1 and T_3 are the delay at f_1 and f_3 respectively, and T_2 the delay at f_2 is given a value of zero.

This method of representing a delay characteristic has a number of advantages. By this representation a given delay may be specified by only two numbers. When a variety of transmission facilities are considered, the quantities M and S will take on a range of values. They may be considered as forming the co-ordinate axes of a plane, on which any facility can be represented as a point. This method of designating transmission circuits can also be applied to delay equalizers which frequently have parabolic characteristics also. In Fig. 3 a number of message facilities and equalizers are located as points on the slope-sag plane. It is usually the case that message circuits are located in the upper half of the plane (positive sag) and equalizers in the lower half (negative sag).

A useful property of the parabolic characteristics is that, if a number of parabolas are added point by point, the sum is also a parabola. In addition, if the slopes of individual parabolas are added, and the sags of individual parabolas are added, the sums are equal respectively to the slope and sag of a parabola resulting from a point-by-point addition of the individual parabolas. Thus, if a given message circuit, designated by a point on the slope-sag diagram, is doubled in length, the resultant delay curve will be a parabola having twice the slope and sag of the original circuit. (For carrier and other facilities where the terminals contribute a large part of the delay distortion, doubling the "length" of the circuit must be construed to include doubling the number of terminals as well.) The point representing the second case would be located on the M-S plane, on the same radial line as the point for the first case, and at twice the distance from the center. Also, if two circuits are connected in tandem, one having values of M and S equal in magnitude and opposite in sign to the respective values of M and S of the other, the result will be cancellation of delay, or equalization. This additive property of the slope and sag quantities makes the task of equalization a rapid and convenient one, since equalization requirements can be prescribed and equalizers can be selected and combined with ease.

This paper gives the results of tests

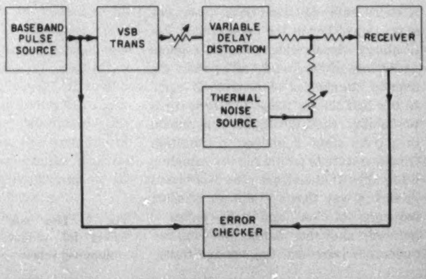


Fig. 4. Experimental setup for slope-sag performance tests

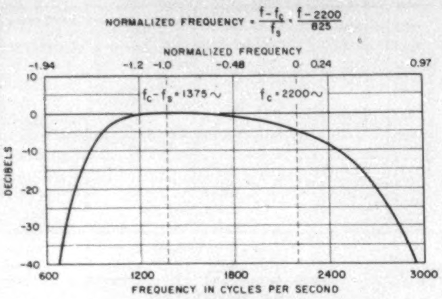


Fig. 5. Transmission characteristic referred to line frequency

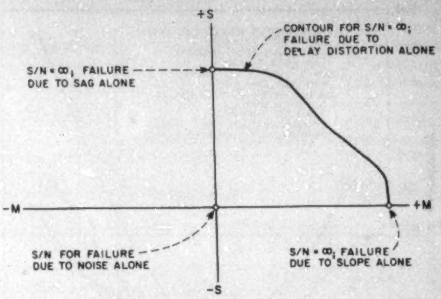

+S
-S
-M
+M
CONTOUR FOR S/N = ∞; FAILURE DUE TO DELAY DISTORTION ALONE
S/N = ∞; FAILURE DUE TO SAG ALONE
S/N FOR FAILURE DUE TO NOISE ALONE
S/N = ∞; FAILURE DUE TO SLOPE ALONE

Fig. 6. Illustration of failure conditions on slope-sag diagram

in which the delay distortion characteristics of transmission systems were considered on a slope-sag basis, and performance of a data system was evaluated in terms of slope and sag. These results provide a means of determining the performance of any facility or combination of facilities having parabolic delay distortion curves when used in connection with a particular data system. Also, the suggestion is made that these results could be expressed on a generalized basis, and may be extended to data systems having a different frequency range, or different bit rates.

Method of Testing

In order to investigate performance on the basis described, tests were made on a data system as shown in Fig. 4. For this system the signal source was a random binary data signal. The individual data pulses have a raised cosine shape. The system operated at a rate of 1,650 bits per second. The data signal was amplitude-modulated on a 2,200-cycle carrier, and was transmitted with a vestigial upper side band.

The performance of the data system was found to be closely related to the average rate of transitions from one state to another in the pulse train. For a binary signal which has a random occurrence of "on" and "off" pulses, the average number of transitions is equal to one half the bit rate. (There is equal probability that the signal will remain in a given state or undergo a transition from one state to another in one sampling interval.) If the signal source is biased in such a way that the number of short sequences of "on" and "off" pulses is reduced, and the number of long sequences is increased, the average transition rate is reduced. For a given signal-to-noise ratio it was found that this type of bias reduced the number of errors or enhanced the performance. If the signal is biased in the opposite manner (more short sequences, the fewer long sequences), the system performance was degraded. For all the tests described in the following discussion, the average rate of transitions was held at one half the bit rate.

The over-all amplitude characteristics of the system including vestigial filtering are shown in Fig. 5. These characteristics were designed with considerable care, in order that the transient response due to amplitude distortion would be minimized. The amplitude characteristics were not changed throughout the test. The over-all delay characteristics with the system connected back-to-back were equalized, except for a small nonparabolic residue which could not be conveniently eliminated.

The objective of the experimental tests was to determine how the data system would perform with a wide variety of different delay-frequency characteristics. A variety of delay curves was obtained by the use of the variable delay network indicated in Fig. 4. It was necessary to use a specially designed network rather than actual transmission facilities, in order to provide control over the shape of the curves, and over the noise applied to the system. With this network, adjustments of delay could be made so that any value of slope or sag of interest to the test could be prescribed.

Thermal noise was introduced into the system through a low-pass filter (having a cutoff at 3,000 cycles), an attenuator, and a mixing pad. The output of the pulse receiver was compared with the transmitter output by the error checker which registered a count whenever a pulse polarity at the receiver differed from what it was at the transmitter. If the number of errors exceeded the error criterion (five errors per minute in this case), the system was considered as "failing"; otherwise it was considered to be "working." Under conditions of no delay distortions, the data system operates at the point $M=0$, $S=0$. The signal-to-noise ratio for this point is the smallest ratio for any point on the slope-sag plane, and for ideal terminals it is equal to the theoretical limit of transmission. This is shown in Fig. 6, as the point for failure due to noise alone.

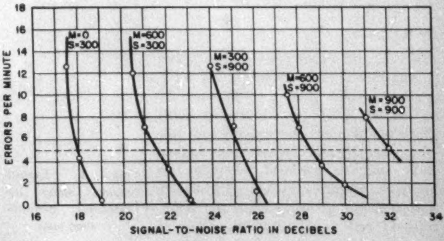


Fig. 7. Error rate curves for various slope-sag values

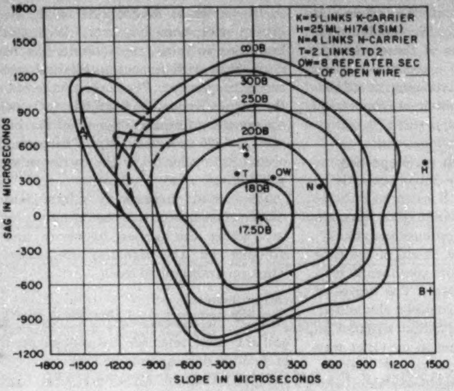


Fig. 8. Slope-sag diagram for vestigial-side-band system, 1,650 bits per second

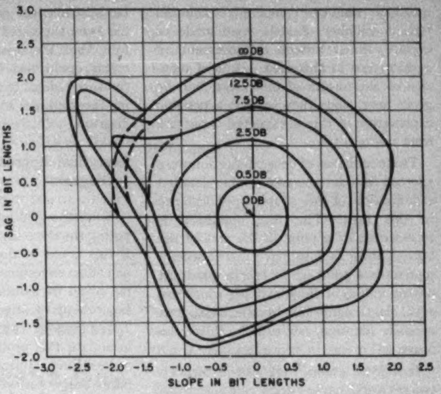


Fig. 9. Normalized slope-sag diagram for vestigial-side-band system

Now consider conditions of no noise, with the slope held at zero. As the sag is increased, a point is found for which the system will fail due to sag alone. Also for conditions of no noise and with the sag held at zero, let the slope increase. There is then a point at which the system will fail due to slope alone. These two points are also shown on Fig. 6. There are many combinations of slope and sag represented by points on the plane, which will also result in failure with the noise removed. If these points are connected, a contour of failure due to distortion alone will result. This is known as the infinite decibel signal-to-noise ratio contour. This contour is shown in the first quadrant in Fig. 6. It may be extended to the other three quadrants as well.

For a point on the *M-S* plane outside of the infinite decibel contour, the data system fails regardless of the noise conditions; therefore, these points were not tested. For points inside this contour the system works with a margin against failure when no noise is present. The margin is determined by the amount of noise which must be added to produce failure.

With the variable delay network set to give a particular slope and sag, the noise was adjusted to a certain level, and the performance in terms of the number of errors was observed for a 5-minute period. By adjusting the noise level to other values and repeating this procedure, a curve of the error rate versus the signal-to-noise ratio was obtained. In Fig. 7 curves of this type are given for five points located in the first quadrant of the slope-sag diagram. The intersection of any of the curves with the dashed line which designates the error criterion determines the signal-to-noise ratio which represents failing conditions for that particular point. Failure in each case is due to the combined effects of noise and delay distortion.

Additional measurements of this type were made for other points ranging over the entire slope-sag plane. A signal-to-noise value was then assigned to each point tested. Many points were found to have very nearly the same margin against failure, and contours of equal signal-to-noise levels were plotted by interpolation. These contours make up what is referred to as the slope-sag diagram shown in Fig. 8.

Results

Several features of the slope-sag diagram should be considered in detail. First it should be noted that the contours of this diagram are related to the characteristics of a particular data system, and to the error criterion used to define a condition of satisfactory performance for that system. If another criterion were used, a second set of contours could also be determined. For a more rigid criterion, the contours will in effect move inward on the diagram, while for a more lenient criterion the contours will move outward.

Also, it is significant to note that in the center of the diagram, the contour lines are spaced relatively far apart. This indicates that when delay distortion is small (slope or sag less than 600 μsec), any change in delay has only a small influence on performance. This means that little is gained by delay distortion correction in this region. With the exception of the *H174* loaded cable, all of the transmission facilities indicated in Fig. 8, fall within 600 μsec of the origin. The contours indicate that these facilities will perform satisfactorily with a signal-to-noise ratio as low as 20 decibels. In the region of relatively large delay distortion (slope or sag greater than 600 μsec), however, the influence of delay distortion becomes more critical, and equalization is more desirable, and even imperative.

The amount and type of equalization necessary to improve a given transmission facility or combination of facilities by a certain amount can now be determined directly from the slope-sag diagram. It is sometimes necessary to establish a compromise type of equalization which will not give complete correction for any one facility, but will allow a number of facilities to operate with the same amount of degradation, or within the same contour on the slope-sag diagram. Equalization of this type can be performed by equalizing for the average position of all the facilities under consideration.

The contours on the slope-sag diagram appear to be circular in shape in the regions of lower distortion. So far this shape has not been confirmed by theoretical analysis. The rather large bulge in the outer contours in the second quadrant cannot be completely accounted for. A number of tests were made to explore the apparent enhancement of performance in this area. A brief mention of the results of these tests will be given here, since they suggest a possible explanation of this unexpected departure from uniformity.

Two conditions of delay distortion were selected in order to make a more detailed examination of the system performance in this area. The first condition is represented by a point in the area of the bulge, shown as A in Fig. 8. The second represents a point in the fourth quadrant having values of slope and sag which are equal in magnitude to the first, but opposite in sign, shown as B in the figure.

Individual pulses were passed through two networks having delay characteristics described by the conditions referred to in the foregoing, and the output pulses were compared. It was found that the transient oscillations in the case of the first network were smaller than those for the second network indicating a confirmation of the reduced degradation in the region of the bulge.

These tests were repeated for two other data systems of the same general type. (One was a system which produced a data signal similar to that described previously. The other was a system having a 2,000-cycle carrier frequency and operating at 1,600 bits per second but otherwise similar to the system tested.) For each of these systems the transient responses were very nearly the same for each network. It was concluded from this that the bulge in the contour pattern was produced by the particular data system used in determining the contour diagram. It was also established that the performance of the data system was very sensitive to minor variations in the delay characteristics when it was operating under conditions of large delay. These facts suggest that the enlargement in the contour pattern may have been due to the unequalized residual in the envelope delay of the data system. Positive confirmation of this explanation could not be made.

It has been shown by E. D. Sunde[1] that for two delay distortion curves, one being the exact negative of the other, the pulse response is reversed in time, but the amount of degradation to the pulse is unchanged. (Equations 2.10, 2.11, and 2.12 of the Sunde paper show that a reversal of sign of the phase characteristics and time leaves the envelope function of a pulse unchanged.) For this condition, the contours of Fig. 8 would be symmetrical about the origin. It has been suggested by Mr. Sunde, therefore, that the results of these measurements could best be compared to other measured data by averaging or adjusting the contours to obtain a symmetrical diagram.

Analytical Approach to Slope-Sag Diagram

Some thought has been given to computing the slope-sag contours analytically. If this could be done, much of the time and effort necessary for experimental testing could be eliminated. The degree of experimental testing could be eliminated. The degree of complication required to formulate this problem in an exact way makes this approach rather difficult, particularly for vestigial-side-band systems as used here. The method of describing the delay distortion curves by a parabola has certain disadvantages from the standpoint of mathematical analysis. This approach was used principally to investigate a wide variety of delay characteristics. As a basis for a computation of pulse distortion, it has not proved to be mathematically tractable. This disadvantage might be overcome by approximating the delay distortion by some other simple shape such as a portion of a sine wave, which could also be specified in terms of two parameters.

A set of slope-sag contours has been computed by P. Mertz by making certain simplifying assumptions. One of these involves a linear weighting of the importance of different portions of the energy spectrum in producing transmission impairment. A weighting was found which gave results somewhat in agreement with the contour determined experimentally, but the analysis indicated that a linear spectrum weighting did not adequately describe the actual conditions.

Application of Results to Other Systems

The foregoing experimental results give a rather complete description of the performance of a particular data system under any condition of parabolic delay distortion. If these results could be made available on a normalized basis, rather than for a specific data system, it is likely that they could be used to express the performance of other systems as well. As a first step in normalizing, these results will be considered to apply only to systems of the same general type as that tested (that is, amplitude-modulated, with a vestigial upper side band) and rule out such systems as double side band, frequency shift, homodyne, etc. With this limitation in mind, these results can be normalized with respect to bit rate and carrier frequency. For the system tested, the carrier frequency $f_c = 2{,}200$ cps, and the signaling frequency (one half the bit rate) $f_s = 825$ cps, which at line frequency occurs 825 cycles below the carrier or at 1,375 cps.

Normalized frequencies referred to carrier transmission may be expressed by subtracting the carrier frequency and dividing by the signaling frequency. Thus, on a normalized basis,

carrier frequency $= 0$
signaling frequency $= -1$ (lower side band)

and the three reference frequencies used in defining slope and sag become

$f_1 = -1.2$

$f_2 = -0.48$

$f_3 = +0.24$

This normalized frequency is shown superimposed on the regular frequency scale, in the transmission characteristics of Fig. 5.

In order to normalize these results, it is also necessary to express the delay, and the slope and sag, in terms of the bit length rather than directly in units of time. When this is done, the slope-sag diagram can be presented in a normalized form, as shown in Fig. 9. This diagram is still restricted to a given type of system with a given error criterion and to parabolic delay distortion. In addition, there are other restrictions on its generality such as a particular amplitude characteristic and method of timing recovery. Its application has been broadened considerably, however, since it can now be applied to many systems having different carrier frequencies and bit rates.

Conclusions

This method of mapping the performance of a data system has proved to be a valuable tool. It has provided an organized approach to a complex problem and results in a clear means of portraying many aspects of a data system in a systematic way. It has increased the general understanding of the effects that various kinds and amounts of delay distortion have on data system performance.

The results of these tests are valid only under conditions of parabolic delay distortion. Since actual delay conditions which depart from the parabola are encountered in practice, there is a need for extending this approach to include other, more complex delay characteristics.

There is a need at the present time for

the results of tests on many types of data systems to be accumulated and compared. This would allow a more complete evaluation of different data systems, and their performance over different transmission facilities. The method of portraying system performance given here could be used as a basis of this comparison.

These results have been generalized to some extent through normalization with respect to carrier frequency and signaling frequency. Through this means they may be used to give an estimation of the performance of other systems of the same general type. There is the possibility that the results may be extended to include additional types of systems as knowledge in this area increases.

The tests described here were time-consuming because of the complexity of adjustments required for determining performance at each point tested on the slope-sag plane. This work could be reduced considerably by the use of specially designed delay networks which could be quickly adjusted to give any required value of slope and sag. This approach to further testing is now being investigated.

Reference

1. Theoretical Fundamentals of Pulse Transmission—I, E. D. Sunde. *Bell System Technical Journal*, New York, N. Y., vol. 33, May 1954, p. 733.

Dial Telephone Service for Smith Island, an Isolated Community in Chesapeake Bay

M. E. LITTLETON
ASSOCIATE MEMBER AIEE

THE utilization of very-high-frequency (vhf) and ultrahigh-frequency (uhf) radio facilities to provide short-haul toll, rural subscriber, or other point-to-point communication systems has answered the demands for modern telephone facilities in small isolated locations. In October 1957, The Chesapeake and Potomac Telephone Company of Maryland combined two radio-carrier terminals and associated central office equipment to provide three communities on Smith Island, Md., as shown on Fig. 1, with the latest dial telephone service.

The present system replaces a unique vhf radiotelephone communication system developed by Bell Telephone Laboratories. This system was installed in 1941 and provided facilities to serve Smith Island and Tangier Island from Crisfield, Md., a distance of 10 and 13 miles, respectively. The principal portion of the radio path is over water and a small portion over marshes and lowlands resulting in line-of-sight transmission. Since the volume of telephone traffic was expected to be small, one radio channel was provided and was shared by both islands. The radio equipment consisted of 160-mc (megacycle) 15-watt amplitude-modulated transmitters and associated receivers located in Crisfield serving both locations on a shared basis. This installation made provisions for six telephones on each island. The Crisfield radio transmitter and receiver were mounted on 50-foot poles with wire extensions to the central office manual switchboard. Although the normal communication was between the islands and Crisfield, it was possible to communicate from island to island by passing through the Crisfield radio equipment and being retransmitted through a by-pass connection. Each subscriber's telephone was equipped with a key that enabled the subscriber to select either a local party line, which furnished intracommunity communications, or a radiotelephone line connecting to the radio terminal. Each subscriber's set had a bell which could be rung by the operator in Crisfield using a selective signaling system.

In 1942, a 9-station magneto private line was provided to interconnect the three communities on Smith Island but did not connect to the radio. While this service provided telephone facilities for the intra-island communication, it did not allow two of the three communities to have direct telephone contact with Crisfield and other mainland points.

In recent years, the demands placed on this service indicated that more adequate facilities were needed. A comparative study was made to determine the most practical method of providing modern telephone communications to span the 10 miles of water separating Crisfield and Smith Island. This study showed that a dial central office with trunks to Crisfield on a multiplexed radio system would be more economical than either a submarine subscriber cable or a dial central office with a submarine trunk cable. There also was the possibility that submarine cable might be damaged during dredging operations in the oyster beds located off the shores of Smith Island. A uhf radio system was selected rather than a microwave system due to the relatively small number of voice circuits required.

Two uhf radio systems now connect Smith Island with the Maryland mainland at Crisfield. Preliminary surveys of the 213 families and 18 business establishments on the island indicated that facilities which would provide a total of six communication channels for off-the-island communications should be sufficient to satisfy local demands for service.

The largest of the three island communities is Ewell and its location, on the northern end of the island, provided not only a favorable location for a central office but also an excellent antenna site. Rhodes Point connects to the new central office at Ewell over 1.5 miles of aerial cable facilities. The southernmost community is Tylerton which is separated from the island mainland by Tyler Creek, see Fig. 2, and is served by aerial cable and 300 feet of submarine cable.

Radio Equipment

Two Dumont *MCA 484B* broad-band radio terminals as shown in Fig. 3, operating in the 450-mc frequency range, provide the radio facilities. The base-band frequency response of 0.2 to 11.2 kc allows for the multiplexing of a voice-frequency channel, two single-side-band suppressed carrier channels together with three associated signaling channels. The broadbanding of this equipment, which had previously been available on a single-channel basis, was accomplished through

---

Paper 58-1292, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 13, 1958; made available for printing September 9, 1958.

M. E. Littleton is with The Chesapeake and Potomac Telephone Company of Maryland, Baltimore, Md.

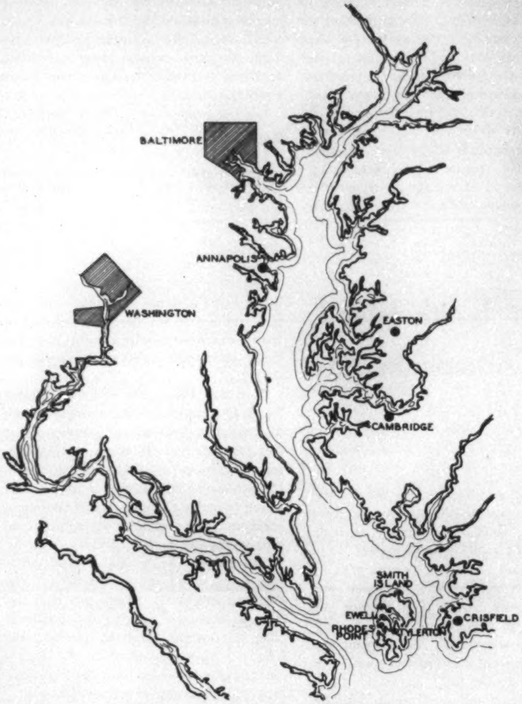

BALTIMORE
ANNAPOLIS
WASHINGTON
EASTON
CAMBRIDGE
SMITH ISLAND
EWELL
RHODES POINT
TYLERTON
CRISFIELD

Fig. 1. Map showing location of Smith Island, Md.

Fig. 2. Submarine cable crossing at Tyler Creek. In the background is typical Smith Island terrain which made the placing of aerial cable difficult

Fig. 3. Two radio transmitter-receiver bays at Crisfield terminal

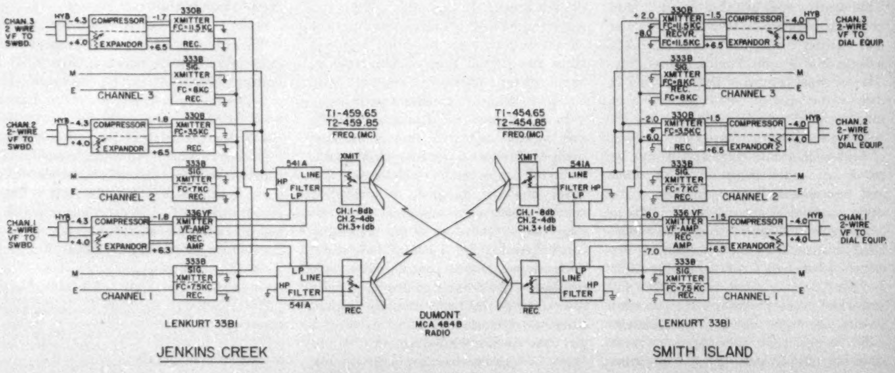

CHAN.3 2 WIRE VF TO SWBD.
HYB
-4.3
+4.0
COMPRESSOR
EXPANDOR
-1.7
+6.5
330B
XMITTER FC=11.5KC
REC.
333B
SIG. XMITTER
FC=8KC REC.
M
E
CHANNEL 3
CHAN.2 2-WIRE VF TO SWBD.
-1.8
330B
XMITTER FC=3.5KC
REC.
333B
SIG. XMITTER
FC=7KC REC.
CHANNEL 2
CHAN.1 2-WIRE VF TO SWBD.
+6.3
336VF
XMITTER VF-AMP.
REC. AMP
333B
SIG. XMITTER
FC=7.5KC REC.
CHANNEL 1
541A
HP
LINE
FILTER
LP
XMIT
REC.
T1-459.65
T2-459.85
FREQ.(MC)
CH.1-8db
CH.2-4db
CH.3+1db
LENKURT 33B1
JENKINS CREEK
DUMONT MCA 484B RADIO
T1-454.65
T2-454.85
FREQ.(MC)
+2.0
-8.0
330B
XMITTER FC=11.5KC
RECVR. FC=11.5KC
-1.5
+6.5
COMPRESSOR
EXPANDOR
-4.0
+4.0
HYB
CHAN.3 2-WIRE VF TO DIAL EQUIP.
333B
SIG. XMITTER FC=8KC
REC. FC=8KC
CHANNEL 3
-2.0
-6.0
330B
XMITTER FC=3.5KC
REC.
CHAN.2 2-WIRE VF TO DIAL EQUIP.
333B
SIG. XMITTER
FC=7KC REC.
CHANNEL 2
-8.0
-7.0
336 VF
XMITTER VF-AMP
REC. AMP
CHAN.1 2-WIRE VF TO DIAL EQUIP.
333B
SIG. XMITTER
FC=7.5KC REC.
CHANNEL 1
LENKURT 33B1
SMITH ISLAND

Fig. 4. Block diagram of radio-carrier system with transmission levels. All levels are in decibels referred to 1 milliwatt with 1,000-cycle tone at the 0 transmission level point

Vo
Pt.
19

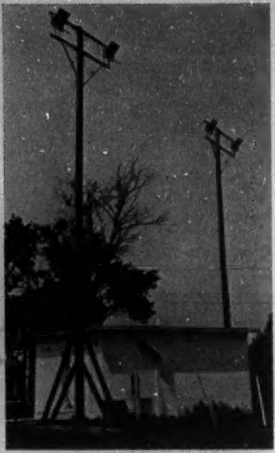


Fig. 5 (left). Smith Island dial central office building and antenna structures

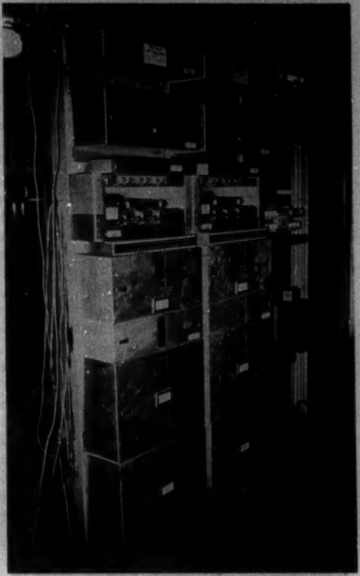


Fig. 6 (right). Two Lenkurt 33B1 telephone carrier systems as installed in Smith Island central dial office

the mutual co-operation of the Dumont Laboratories and the Bell Telephone Laboratories. For continuous duplex operation, the operating frequencies of the transmitter and receiver are separated by 5 mc. Actual operating frequencies as well as transmitted levels are shown in Fig. 4.

Antennas

Each radio terminal utilizes four Andrew's hurricane-type corner reflectors mounted as shown in Fig. 5. The supporting poles are 65 feet high with a 30-foot separation between bases. Special footings were needed at both locations to prevent excessive settling of the antenna poles into the soft terrain while reinforced guying was provided as protection against the ever-impending bay storms. The transmitting and receiving antennas of each radio system are horizontally polarized and mounted on a common pole. They are spaced 10 feet apart to provide the required decoupling between the transmitting and receiving paths. The antenna poles of the two radio system are separated by 30 feet to provide isolation between the two radio systems.

Multiplex Equipment

Two Lenkurt type *33B1* telephone carrier systems, as shown in Fig. 6, provide a compact arrangement for the application of three telephone channels to broad-band radio communications facilities. Each carrier system will provide one voice frequency, two carrier-derived channels, and three signaling channels. The voice-frequency channels occupy a band of frequencies between 200 and 3,000 cycles per second. Two additional channels are carrier-derived and convert voice frequencies of 300 to 2,700 cycles per second to single-side-band suppressed-carrier amplitude-modulated transmission at higher frequencies. Channel 2 transmits and receives on the upper sideband of 3.5 kc and Channel 3 transmits and receives on the lower sideband of 11.5 kc, as shown in Fig. 7. The block diagram of three channels of the radio carrier system including compandors and voice-frequency terminating units are shown in Fig. 4. Signaling for dialing and supervision on the three channels is transmitted and received at carrier frequencies of 7.5, 7.0, and 8 kc for Channels 1, 2, and 3, respectively. Each channel utilizes a compandor unit[1] to provide an effective increase in signal-to-noise ratio. Two-wire facilities connect the Crisfield central office to the radio-carrier equipment located at the Crisfield terminal building.

Dial Central Office

The new office established on Smith Island is a Western Electric type *356A* step-by-step community dial central office with a capacity of 200 individual lines. The office is arranged for normal 2-letter 5-digit dialing and is identified as HAZEL-5. While local intra-island calling is made on the 2-letter 5-digit basis, incoming and outgoing calls to Crisfield and distant points must be routed through an intermediate operator in the Crisfield central office. Island subscribers may reach the Crisfield operator by dialing 0. With the advent of direct-distance dialing in this area, no major modifications to the Smith Island central office equipment will be required. Direct trunks will then be provided to the toll center located at Salisbury, Md., utilizing a combination of radio and land carrier systems. The

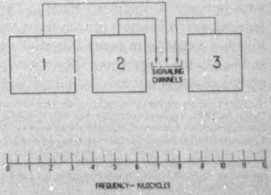


Fig. 7. Multiplex channel allocations

Smith Island–Crisfield trunks, described herein, will then provide for direct dial service between the two communities.

Alarms and Controls

The alarms and controls of this communications system appear as both audible and visual alarms on the Crisfield cental office switchboard with instructions to authorized personnel for their interpretation and operation.

The functions of the Crisfield radio and carrier equipment are transmitted over regular telephone cable facilities.

The remote indication and control of the Smith Island equipment is obtained by the integration of the built-in radio-carrier functions and the regular alarm system associated with the dial central-office equipment. In the event of an equipment failure on Smith Island, the appropriate alarm relay will seize an idle voice channel and will light a lamp associated with that trunk on the Crisfield switchboard. Upon receiving no response on this trunk, the operator will know that a trouble condition exists. The operator will then dial a predetermined level code into another Smith Island trunk to determine the nature of the alarm. "No tone" will constitute a major alarm, a "busy signal" indicates a minor alarm, while a satisfactory condition is indicated by two rings. Alarm indications include low transmitter power, low received sensitivity, emergency a-c power failure, and several types of central office equipment failures.

The Smith Island transmitters are controlled from Crisfield by dialing a level code into an idle radio trunk which when received in the island central office will turn the appropriate transmitter on or off.

Power

The basic power for the radio-carrier system is single-phase 115-volt a-c, 60-cycle. An interesting side note on the power for Smith Island is the fact that commercial power is generated on Tangier Island using the diesel-driven generator facilities of the Chesapeake Island Electric Cooperative. The power is then fed over 10 miles of high-voltage lines on poles located in the intervening marshes with high level crossings over navigable channels.

The emergency a-c power for the buildings on Smith Island and Crisfield is provided by Koehler diesel-driven automatically controlled alternators with a single-phase 60-cycle output of 10 kw at 230/115 volts a-c. In the event of a power failure the generator will start automatically after a 3-minute time delay and will continue to operate for 5 minutes after the initial restoration of power.

Conclusions

The use of two radio-carrier systems in conjunction with a dial central office has provided the isolated communities on Smith Island with a modern, reliable telephone service. The acceptance of this service by the islanders is evidenced by the fact that the number of telephones has increased from 120 to 140 in a 4-month period.

Reference

1. A Miniature Compandor for General Use in Wire and Radio Communication Systems, F. S. Boxall, R. S. Caruthers. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 804–11.

Public Air-Ground Telephone Service Trial

L. M. AUGUSTUS
NONMEMBER AIEE

WHILE radio communication to airplanes has been in everyday use for many years, until September 15, 1957, the only public telephone service available to aircraft has been that furnished through coast stations which have been established for the provision of telephone service to ships. Approximately 600 aircraft are registered for service through coast stations operating in the 2-mc (megacycle) frequency range. The service is not satisfactory in many respects. It is not available in many inland areas and due to the crowded conditions on the marine channels fast-moving aircraft are frequently unable to receive any service before landing at their destination. In addition, atmospheric noise at these frequencies seriously limits the range and reliability of the service. Even with these limitations the service has grown as shown in Fig. 1.

In order to obtain practical experience with the aircraft communications problem, two experiments have been conducted. At the time of the Provisional International Civil Aviation Organization meeting in Indianapolis, Ind., in October 1946 in co-operation with the Civil Aeronautics Authority (CAA), an airplane was equipped for operating with urban mobile stations in the 152–162-mc range. This installation was used experimentally during the meeting for communication through the Indianapolis urban mobile land station facilities which were, at the time, undergoing test. In addition to the experiments with the CAA airplane, a United Air Lines Douglas *DC-6* airplane was fitted with 152–162-mc land mobile equipment and experimental calls were made during the inaugural flight and certain other flights thereafter. The tests indicated that an aircraft flying at altitudes from 10 to 20,000 feet would cause interference with land mobile stations using the same channel within a radius of 200 miles or more. The experiments showed conclusively that a good grade of telephone service could be given to aircraft employing frequencies in the 150-mc

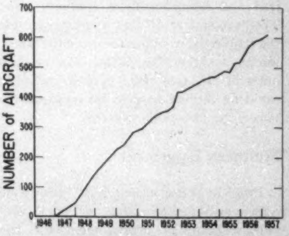


Fig. 1. Growth of aircraft radiotelephone service shown by aircraft equipped to use coastal harbor service

---

Paper 58-1293, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing September 9, 1958.

L. M. Augustus is with the Michigan Bell Telephone Company, Detroit, Mich.

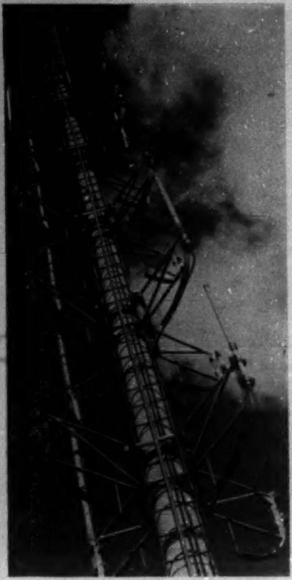


Fig. 2. Antenna installation at Chicago

range but that channels separate from those employed for land mobile telephone service would have to be used.

The growth in the number of civil aircraft has been phenomenal since the end of World War II. According to current estimates a total of about 65,000 active civil aircraft are flying more than 13 million hours a year. Several aircraft-user groups have repeatedly expressed a growing interest in public air–ground radiotelephone service connecting to the land-telephone network. This was specifically expressed by the National Association of State Aviation Officials in a resolution adopted by the Board of Directors in Washington, D.C., on February 2, 1956, and submitted to the Federal Communications Commission (FCC). This resolution together with a request from the National Business Aircraft Association for a "Special Service Communications Channel" was subsequently considered by Special Committee *56* of the Radio Technical Commission for Aeronautics. Since that time several informal meetings have been attended by representatives of commercial airlines, military services, operators of executive and private aircraft, radio equipment manufacturers, and telephone companies. The interest expressed at these meetings indicates that there is a compelling need for air–ground telephone service connected with the land-line telephone network.

Early in 1956 a government–industry meeting was held in Washington under the chairmanship of C. A. Petry of Aeronautical Radio, Inc., to discuss "in-flight radiotelephone service to aircraft."

At subsequent meetings with representatives of manufacturing companies interested in building air-borne radio equipment, the U. S. Independent Telephone Association, Aeronautical Radio, Inc., and the Bell System, it was generally recognized that the telephone industry is in the best position to provide the ground facilities for public telephone service to airplanes. In order to obtain suitable frequency allocations to provide the numbers of channels such a service may need within a few years, it is necessary to develop plans for service on a nationwide basis.

Trial

As an aid in evaluating some of the questions encountered in developing such plans, a limited trial has been undertaken with ground stations at Detroit, Mich., and Chicago, Ill. The purpose of the trial is to provide needed technical data and traffic operating experience under actual flight conditions over a substantial part of an important route.

The frequencies used are 454.95 and 459.95 mc, normally assigned to the Domestic Public Land Mobile Service. These frequencies were chosen in order to obtain transmission information for frequencies well above those presently used for commercial operational air–ground communication. Theoretical consideration of path loss and antenna requirements lead to the conclusion that frequencies no higher than those in the lower part of the ultrahigh-frequency band will be feasible without appreciably restricting the range. These are also the highest frequencies for which commercial mobile equipment is readily available. Under a developmental grant the FCC has authorized the use of these frequencies at Chicago and Detroit for service to aircraft until September 1959. Fig. 2 is a view of the antenna installation on the Field Building in Chicago. Fig. 3 is a view of the Detroit antenna installation at Dixboro, Mich. The Detroit site is in relatively flat open country. The Chicago antenna installation is approximately 535 feet above the surrounding land while the Detroit antenna height is approximately 50 feet.

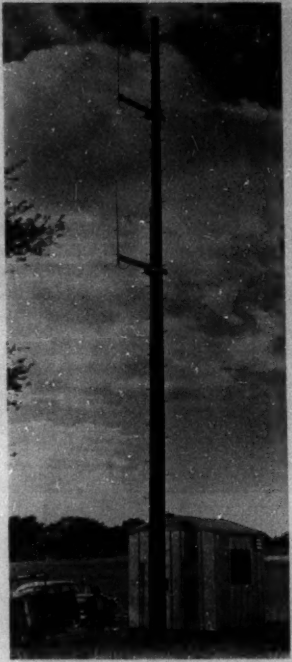


Fig. 3. Detroit transmitting and receiving station

Tests at these widely different sites should provide useful information for most of the base stations which would normally be encountered for a nationwide service. Frequency modulation was employed for these tests in view of its demonstrated advantages for a service to be used by the public. One of these advantages, the capture effect, appears to be of unusual significance in air–ground service where cochannel interference is a principal problem. The capture effect can be illustrated by analysis of a tape recording taken in a *DC-3* plane traveling 180 mph at an altitude of 8,000 feet on a direct route from Detroit to Chicago with transmitters at both these cities operating on the same frequency, modulated by different talkers. At the beginning, while the plane is in the area where the Detroit station is controlling, only the Detroit talker is heard. As the plane moves into the interference zone, a heterodyne appears, caused by the Chicago station. The impairment to the Detroit transmission is slight for a period of about a

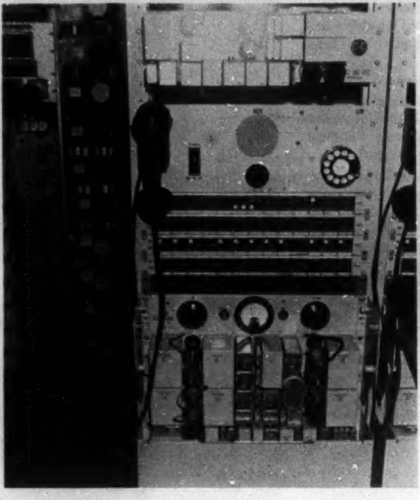


Fig. 4 (left). Western Electric G-2 control terminal

Fig. 5 (above). Switchboard positions at Detroit station

minute and a half, then rapidly becomes worse until the intelligence is completely lost due to the near-equality of the two received signals. The transition period lasts about a half-minute, before the Chicago talker is heard identifying the Chicago station, with the Detroit talker just apparent in the background.

With a ground speed of 180 mph, the half-minute transition represents $1^1/_2$ miles of flight. Other recordings taken at higher altitudes indicate a more extensive zone of interference as might be expected from the greater overlap as altitude is increased.

Since the coverage areas of the two stations overlap considerably at high altitudes, the operating practice will be such that the two stations will not transmit simultaneously if either is working an aircraft in the interference zone. By an arrangement including a co-ordinating circuit and monitoring facilities each operator will know when the other base station is handling traffic and whether the aircraft using the channel is in the overlap zone. Under these conditions, she will not energize the local transmitter.

Transmitting and Receiving Equipment

The equipment arrangements in the Detroit station used in this system were designed by the Bell Telephone Laboratories to incorporate the features which have been found to be desirable for common-carrier mobile radio services over the past 10 years. For example, a radio-frequency indicating device coupled at the transmitter output is used to provide an indication to the traffic operator that the transmitter is energized and she may dial the called party. This same feature will also indicate, by automatic alarms to the technical attendant, a failure of the transmitter or a false operation, should the transmitter radiate for any reason other than a signal from the remote control equipment.

The receiver is conventional in design: double-conversion superheterodyne with two grounded-grid radio-frequency stages. The electronic squelch of the receiver has been disabled and the normal noise voltages developed in the receiver are used to control a carrier-operated relay. The relay contacts are connected in such a manner that an idle receiver has a ground applied on the grid of the audio output stage. When a carrier is received and the receiver quiets, this ground is removed and transferred to a control circuit which indicates to the remote control equipment that a signal is being received. The recovered modulation of the incoming signal is then allowed to pass through the receiver and ultimately to the aviation service operator.

Vertically polarized omnidirectional gain antennas were used at both Chicago and Detroit. At the Detroit station, arrangements were provided to permit observation of the effect of small changes in the height and to determine the need for height diversity. This can also be seen from Fig. 3.

The transmitters used at both stations were phase-modulated, having a nominal power output of 15 watts with very good frequency stability. The use of cavity filters in a diplexing arrangement enables the operation of the transmitter and receiver simultaneously from a common antenna.

Control Equipment

In planning for this new service it was decided that the system arrangements should be the same, at least in the beginning, as the existing land mobile common-carrier services. This permits utilization of the Western Electric *G2* control terminal, shown in Fig. 4, which has proved very satisfactory as the intermediate connecting apparatus between the radio equipment and the "Aviation Service" operator, who supplies the final connection to the land telephone. The operator positions are shown on Fig. 5.

The control terminal does many things to simplify the use of radio circuits by the traffic operators, but its principal functions are to convert 4-wire to 2-wire operation, supply signaling tones for dialing the mobile station, and maintain a relatively constant audio volume to modulate the transmitter. These are discussed briefly as follows:

1. The standard land telephone is connected to the central office, and ultimately to any other telephone in the nation-wide network by means of a single pair of wires

Vol.
Pt. 1
1958

Fig. 6. Control unit and noise-canceling handset

over which the user both talks and listens. Contrasted with this, the radio circuit has separate talking and listening equipments which break the communication path into two one-way circuits, called 4-wire operation. In the control terminal, the talking and listening paths are bridged to the 2-wire line by a resistance network which delivers the proper audio level from the receiver to the switchboard and also to the transmitting path for modulation of the base transmitter. By modulating the base transmitter with the received audio the station can function as a repeater and furnish mobile-to-mobile communication over the coverage area of the base station.

2. The signaling method adopted by the Bell System utilizes a 2-tone sequential system using the frequencies 600 and 1,500 cycles per second. The arrangement is such that the tones are sent alternately, each transition of tone resulting from a pulse of the operator's dialing equipment. Since the number of pulses corresponds to the value of the digit dialed, the number of tone transitions is equal to the dialed digit. The mobile selector will advance one step for each transition of tone. If the selector is set up properly, a signaling contact will close when the dialing is completed. Five-digit telephone numbers will be assigned, the sum of the digits being equal to 23, 24, 25, or 26, with the digit 1 not used, and zero equal to 10. The digit 1 is not assigned because it is automatically transmitted as a clearing signal to the mobile receivers to set all selectors to the start position at the beginning and end of the dialing period.

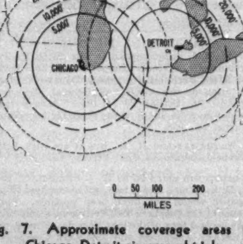


Fig. 7. Approximate coverage areas for Chicago–Detroit air–ground trials

3. Due to the nature of the telephone plant and telephone users, there is the additional problem of attempting to provide suitable modulation of the transmitter on every call from any telephone. The range of audio levels received at the aviation switchboard from land subscribers is expected to be from 0 volume units to −40 volume units, the distribution following a normal distribution curve. To compress this range of volumes to values which will allow good modulation on the great majority of transmissions, the terminal includes a self-adjusting variable-gain amplifier which will deliver zero ±4 volume units to the transmitting cable pair.

The electrical noise interference on board the *DC-3* was found to be inconsequential at these frequencies. The high ambient acoustic noise in the plane was largely overcome by the use of a noise-canceling microphone. The use of a conventional handset receiver as shown in Fig. 6 permitted satisfactory reception over the high airplane noise.

Results

A large volume of propagation data has been obtained by recording the field strength from the plane while in flight along various radials and at several different altitudes. Analysis of these data have not been completed at this time, but it appears with the powers and antennas used, that coverage is confined to regions lying well above the radio horizon which is the optical horizon modified by a factor of 4/3 to account for average refraction in the atmosphere. For example, with the 50 foot antenna at Detroit, the radio horizon to a plane at 10,000 feet is 150 miles. In actual practice the range obtained over relatively flat terrain was some 110 to 125 miles. Along a radial where the terrain fell off toward the lake in a manner to add to the effective height of the Detroit station, the range was some 25 miles greater. With the 535-foot base station antenna at Chicago ranges approaching the radio horizon were obtained. To a plane flying at 10,000 feet the radio horizon is approximately 175 miles. This is indicated on Fig. 7, which shows the approximate coverage areas of the Chicago and Detroit stations to aircraft at altitudes of 5,000, 10,000, and 20,000 feet.

Fig. 8. Location of antenna on DC-3 test aircraft

At the coverage limits the service was found to cut off in an abrupt manner. Particular attention was given to lobing effects caused by interaction of direct and ground reflected waves as has been so adequately described in the papers by Bureau of Standards personnel.[1] It appears that the lobing structure is less well defined than theoretical considerations would indicate, due presumably to the presence of multiple reflections from terrain irregularities. Lobes were found to be more numerous from the Chicago station using the 535-foot antenna as would be expected from theoretical considerations.

The antenna installation on the General Motors Corporation *DC-3* as shown in Fig. 8 resulted in a nonuniform pattern which means that the coverage range will depend somewhat on plane orientation.

When the plane flew through the region directly above the ground station, wide excursions of field strength were experienced but no "cone of silence" could be observed.

Problems

Development of an adequate nationwide air-ground service will require the solution of several outstanding problems. As has been touched on previously, allocation of a suitable band of frequencies is a prerequisite. An air-ground system using known techniques could be developed to operate in almost any band below 1,000 mc. However, use of frequencies between 100 and 500 mc would offer important advantages over higher frequencies such as earlier availability of equipment, lower cost, and better radio-propagation characteristics. It is estimated that a public radiotelephone service for aircraft will require 40-50 channels on a frequency-modulation basis to provide for the traffic which will develop in the next 8-10 years.

As was mentioned earlier, the present service is operating under a developmental grant from the FCC and, without further authorization from the Commission, service must end in September 1959. The existing trial has received the enthusiastic support of the entire aviation industry. An immediate need exists for the present service not only to be continued, but extended to cover the more important air lanes. The Bell System, therefore, on March 31, 1958, petitioned the FCC as a first step to extend the service to the east. The service would then cover cities originating one third of all commercial air travel in the United States and more than one fourth of all private flights. The addition of three stations near Pittsburgh, Pa., Washington, and New York, N. Y., should provide a coverage corridor 200 miles wide extending from 100 miles west of Chicago to the East Coast.

High-flying aircraft have a greater transmission range than do low-flying aircraft. For example, the range as limited by the radio horizon is less than 100 miles for a plane flying at 4,000 feet and may be in excess of 250 miles for a plane flying at 40,000 feet. It is evident that any system engineered to provide coverage for a particular altitude would result in the following:

1. Gaps in coverage between base stations for aircraft flying below this altitude.
2. Interference due to multiple reception of base station signals by aircraft flying above this altitude. The cochannel interference problem demonstrated by the tape recording can be considerably more severe unless careful attention is paid to this factor.

One problem which must be considered is that of maintaining a continuous conversation to an airplane flying between two base-station coverage areas. Liberal use of extensive telephone plant may be necessary to avoid the inconvenience of reestablishing the connection.

Reference

1. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Apr., May, 1952.

Radio Attenuation at 11 Kmc and Some Implications Affecting Relay System Engineering

S. D. HATHAWAY
NONMEMBER AIEE

H. W. EVANS
NONMEMBER AIEE

INCREASING use of the common-carrier microwave-frequency bands at 4 and 6 kmc (kilomegacycles) has directed attention to the next higher band at 11 kmc. All three bands are subject to atmospheric fading, but propagation at 11 kmc differs from that at the lower frequency bands chiefly in its vulnerability to rain. Knowledge of the statistics of the excess path loss caused by rain is a necessary prerequisite to 11-kmc system design, and therefore an experiment was undertaken to extend the modest body of available knowledge.

The effects of rain on microwave radio propagation have been calculated by Ryde and Ryde.[1,2] The radio energy is absorbed and scattered by the rain drops, and these effects become more pronounced at the higher microwave frequencies where the wave length and the raindrop diameter become more nearly comparable.

The excess attenuation caused by rainfall depends on the number of drops per unit volume in the radio path, the square of the drop diameter, and a complex factor representing the ratio of the total energy absorbed and scattered by a single drop to the energy in that area of the wavefront equal to the projected area of the drop.

Laws and Parsons[3] observed the distributions of drop sizes for various rates of fall on a horizontal surface, using the method of Bentley.[4] (The Bentley method involves exposing trays of sifted flour to the rainfall, baking the flour to solidify the pellets formed by impinging raindrops, and then sorting the pellets by size. The flour has been calibrated by generating drops of known size, so that drop size can be determined from the pellet size.) The higher the rainfall rate, the larger the drops, and also the greater the spread in size of drops. Ryde computes the number of drops per unit volume from the data of Laws and Parsons by applying the terminal velocity appropriate to the drop mass.

The excess path loss per mile according to Ryde for the three common-carrier frequency bands of 4, 6, and 11 kmc, is shown in Fig. 1 for various rates of rainfall. (It is interesting to note that from his computations Ryde concludes that the excess attenuation caused by hail is in the order of one-hundredth that caused by rain, that ice crystal clouds cause no sensible excess attenuation, and that snow produces very small attenuation, even at

---

Paper 58-1236, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted June 25, 1958; made available for printing August 26, 1958.

S. D. HATHAWAY and H. W. EVANS are with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The authors are indebted to K. J. Frolund, G. H. Swan, D. K. White, and others, without whose help this paper would not have been possible. The assistance of the Southern Bell Telephone Company and the central area of the long lines department of the American Telephone and Telegraph Company is deeply appreciated.

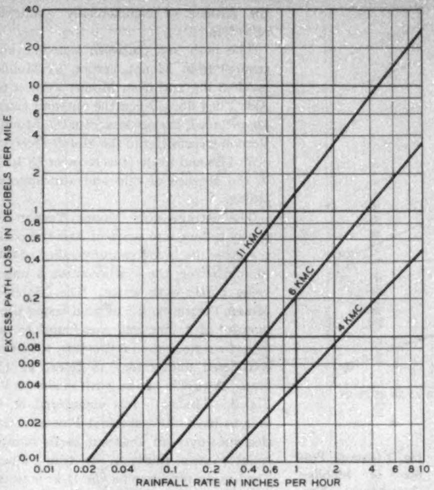


Fig. 1 (left). Rain attenuation versus rainfall rate (theoretical, after Ryde and Ryde)[1]

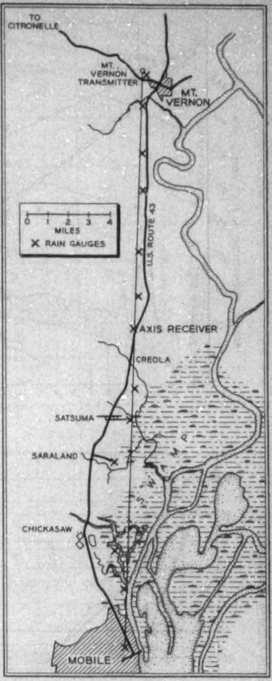


Fig. 2 (right). Map of Mobile–Axis–Mount Vernon radio path

the excessive rate of fall of 5 inches per hour.)

Many rainfall data are available for point locations, but very little is known about the relationship between the rate of fall at a single rain gaging point and the profile of rate of rainfall along a radio path. Furthermore, most rainfall data are in terms of fairly long discrete intervals, such as 30 minutes or one hour, and the relationship between hourly and instantaneous rates of fall is not perfectly known.

Bussey[5] has analyzed rainfall data for one year from the Muskingum River watershed in Ohio,[6] and finds that the annual distribution of 1-hour point rates is approximately the same as an annual distribution of instantaneous 50-kilometer path rates. He further suggests that 10-minute point data may apply to an 8-kilometer path, 30-minute data to a 25-kilometer path, etc.

It was the purpose of the experiment described here to seek confirmation of Ryde's relationship between excess path attenuation and instantaneous rate of rainfall, and to measure the profile of rate of rainfall along a radio path in hopes of finding correlation with rainfall measured at a single point. It was expected that this information would be useful in determining design parameters for 11-kmc radio-relay systems and for suggesting the conditions under which they be used.

The experiment consisted of operating a radio path of a length typical of short-haul radio-relay systems in a heavy-rain area for a year. Instrumentation included devices for measuring excess radio path loss, and rain gages along the path at intervals short enough to define the rainfall profile.

Radio Path

The requirements that determined the choice of the radio path were:

1. Heavy rainfall, both in rate and depth.
2. Length of about 25 miles, which is considered typical of possible 11-kmc application, with the possibility of a second receiver mid-way in the path so that some feel of interpolation versus length would result.
3. All-weather highway parallel to and very near the path, to permit access to rain gages.
4. Existing structures and buildings for antennas and radio equipment.
5. Preferably equipped with an operating 4-kmc radio-relay system, so that some comparison could be made between 4- and 11-kmc propagation.

Literally hundreds of possible paths were examined. The choice narrowed quickly to the Gulf Coast region because of the high incidence of heavy rainfall and the great total rainfall which is in the order of 60 inches per year. (Annual depths of about 150 inches occur in the north Pacific Coast rain forests, but surprisingly the rate of fall is quite low.) Finally a path was selected between Mobile and Mount Vernon, Ala., 27.7 miles long, approximately north and south, and parallel to a good highway, as shown in Fig. 2.

Arrangements were made to locate the transmitter at the Mount Vernon *TD-2* 4-kmc radio-relay station, and an 8- by 12-foot plane reflector, specially made to be flat to 1/16 inch, was mounted at the 300-foot level of the *TD-2* antenna tower. The reflector was illuminated by a 5-foot parabolic antenna on top of the transmitter equipment housing, using a buttonhook feed constructed of commercially available waveguide pieces. The gain of the antenna system at Mount Vernon was 44.2 db (decibels), 1.2 db greater than the gain of the parabolic antenna alone.

At Mobile, a similar antenna system for the receiver was placed on the *TD-2* tower atop the telephone building; but, because this tower was only 85 feet tall, a 6- by 8-foot plane reflector was used with the 5-foot parabolic antenna. The gain of the antenna system at Mobile was 42.8 db, 0.2 db less than that of the parabolic

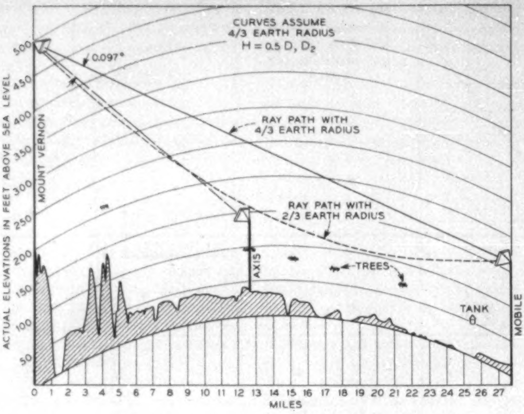


Fig. 3 (above). Profile of Mobile–Axis–Mount Vernon radio path

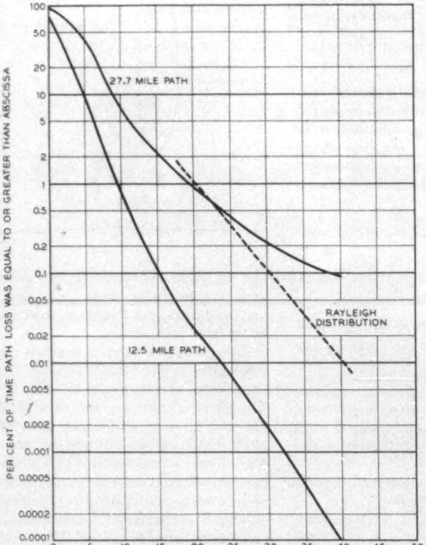


Fig. 4. Distribution of selective fading, March 1, 1956, through July 31, 1956

antenna alone because of the small spacing between the antenna and the reflector.

At a point 12.6 miles south of the transmitter at Mount Vernon, near Axis, Ala., a second receiving station was constructed. It used a 104-foot path-loss testing tower,[7] with a 3-foot parabolic antenna having a gain of 37.1 db mounted directly on the receiver front end, which could be run up and down the tower on a carriage. The tower was located directly in the path from Mount Vernon to Mobile.

The antenna sizes were chosen to produce roughly equal received signal levels at Mobile and Axis, and to produce as large a received signal as was consistent with physical stability of the towers and the flatness of commercially available reflectors.

The path loss (between isotropic antennas) from Mount Vernon to Mobile is 146.3 db, and from Mount Vernon to Axis 139.4 db. When the antenna gains are included, the net loss from the Mount Vernon transmitter to the Mobile receiver is 59.4 db and to the Axis receiver 58.1 db in the absence of rain and atmospheric fading.

The terrain near Mount Vernon is gently rolling, and south of Axis the path traverses the broad swampy valley of the Mobile River, the southernmost 3 miles being partly over water. The Mobile–Mount Vernon path had been tested previously[8] at 4 kmc and was found to be free of strong ground reflections, so it was considered unnecessary to retest at 11 kmc. A profile of the path is shown in Fig. 3. The path was engineered at 4 kmc to have one third first Fresnel zone clearance over an apparent earth radius equal to two thirds of its true value. The passive reflectors for the 11-kmc tests had to be mounted below the existing *TD-2* system antennas for physical reasons. The effect of lower antenna heights was to reduce the path clearance by 15 feet relative to the 4-kmc path. The limiting point in the path occurred 18 miles from the Mount Vernon end and with 60-foot trees the clearance was 36 feet when the effective earth radius was two thirds of its true value. At 11 kmc this was approximately 0.7 first Fresnel zone, which is considered adequate even for this area. The shorter path, Axis–Mount Vernon, had clearance in the order of four Fresnel zones, which was far more than sufficient.

Results

FADING

Fig. 4 shows the signal level distribution of both paths due to multipath fading for a 4-month period, omitting the effects of rain. The long path distribution exceeds Rayleigh for fades greater than 20 db. This would seem to indicate a strong stable reflection condition due either to ground reflections or to layer stratifications in the atmosphere. Since no path loss tests were made at 11 kmc, ground reflections cannot be ruled out. However, the path for the most part traverses land covered with low vegetation and pine forests, usually thought to be nonreflective. The *TD-2* system suffered similar fading, and it had been established that at 4 kmc the path was essentially nonreflective. Unfortunately, only a slow-speed strip recorder was moni-

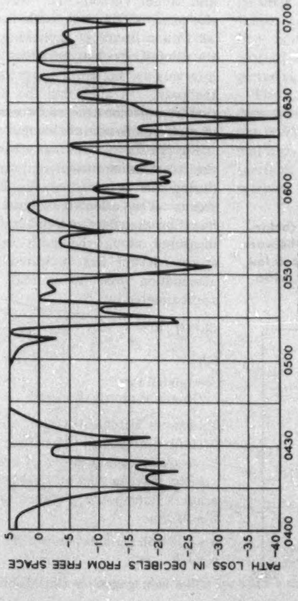


Fig. 5. Typical selective fading at 4 kmc

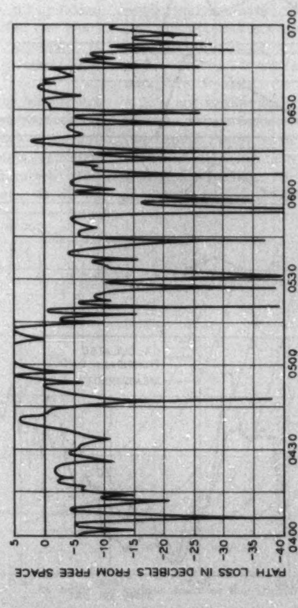


Fig. 6. Typical selective fading at 11 kmc

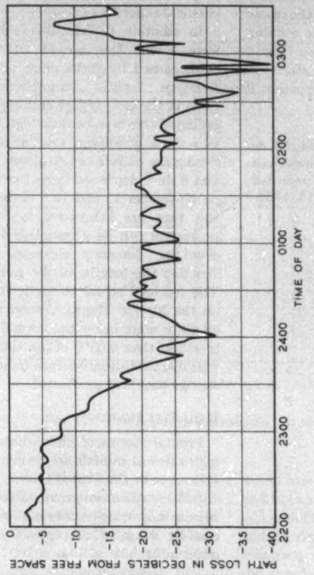


Fig. 7. Earth bulge fading at 4 kmc

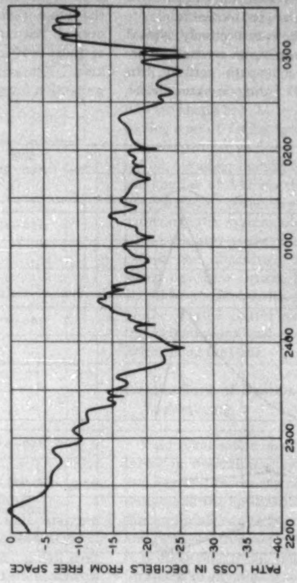


Fig. 8. Earth bulge fading at 11 kmc

toring the 4-kmc system and comparative distribution data are not available.

Figs. 5 and 6 show, respectively, typical 4-kmc and 11-kmc signal strengths during periods of multipath activity. In general, multipath fading on 4 kmc and 11 kmc began and ended at approximately the same time and followed the same over-all pattern. However, the number of fades was greater at 11 kmc than 4 kmc. The necessity of having diversity protection for such systems is apparent if they are to meet long-distance telephone circuit standards.

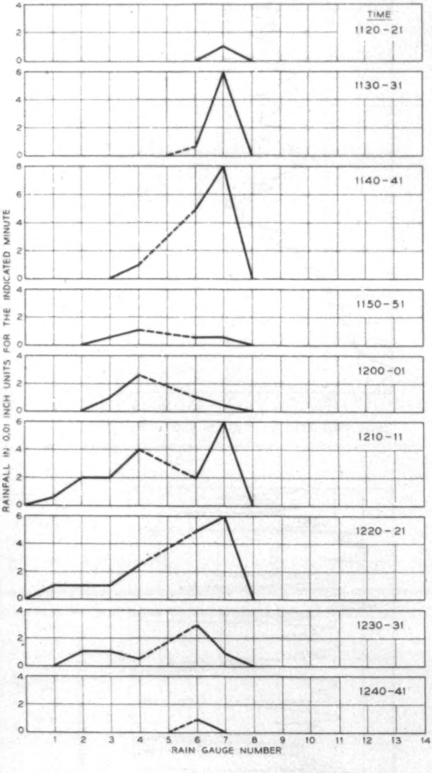


Fig. 9. Rainfall distribution over Axis–Mount Vernon path, March 15, 1956

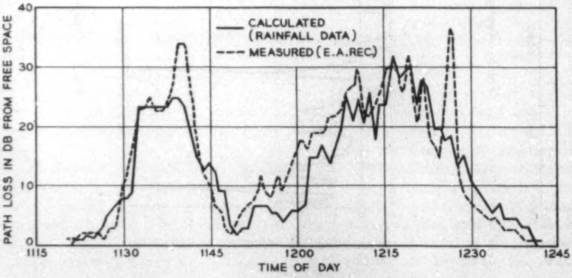


Fig. 10 (below). Correlation between rainfall and path loss, March 15, 1956

In addition to selective fading there were several long periods of depressed fields caused by earth bulge or obstructive-type fading. Atmospheric conditions in the Gulf region are favorable for fading of this type because high humidity and stable conditions exist at night and during the early morning hours. Figs. 7 and 8 show depressed fields that occurred on November 18 and 19. Since fades of this type are insensitive to frequency, protection can be accomplished only by providing adequate clearance and restricting the lengths of the radio paths. The received signal strength at 11 kmc on the Mobile–Mount Vernon path was 40 db or more below the normal received level less than 0.03% of the time for the year due to obstructive-type fading. The shorter path was unaffected.

RAIN ATTENUATION

Frontal storms of short duration and high rates of rainfall are common in the delta area of the United States. In general, these storms originate in the Gulf of Mexico and travel landward in a northeasterly direction. A typical storm arrived in Mobile on March 15, 1956, and passed diagonally across the radio test path between the small towns of Creola and Mount Vernon, Ala. Rain fell over this area of the test path from about 11:15 a.m. to 12:45 p.m. Fig. 9 shows the rainfall rate distribution at 10-minute intervals as the storm progressed across the path. In analyzing the data, such profiles were constructed for each minute of each significant rain event.

Fig. 10 shows the correlation between the measured and calculated signal levels during the progress of the March 15 storm. The calculated signal level is based on the effective 2-mile rainfall rate measured along the path during the storm. Ryde[2] has indicated that the attenuation due to rainfall can be approximated by

$$\text{db} = k\int_0^r R^{\alpha} dr$$

where

R = rainfall rate
r = length of propagation path

Professors Hitchfeld, Gunn, and East of McGill University, Montreal, Canada, have made computations of k and α for various wavelengths at 18 degrees centigrade. Extrapolation of their work yields for 11 kmc:

$k = 1.395$ db per mile per inch per hour
$\alpha = 1.3$

The rain gages on the Mobile–Mount

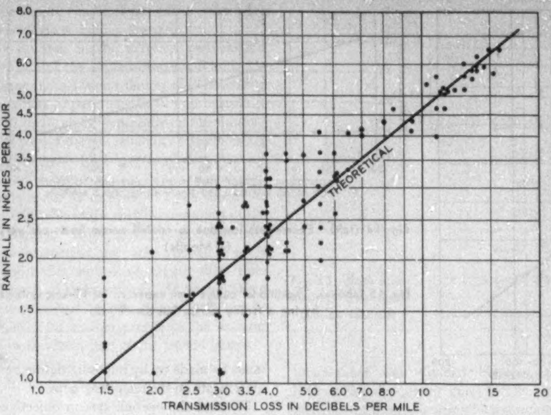


Fig. 11. Scatter plot of transmission loss versus rainfall rate

Vernon path were placed at approximately 2-mile intervals. The rainfall rate within one mile either side of a rain gage has been assumed uniform. Then, over any 2-mile path the attenuation due to rainfall is approximated by

$$\text{db} = 2kR^{\alpha}$$

The attenuation over the entire path is then the addition of the attenuations due to the 2-mile segments of the path:

$$\text{db} = 2k(R_1^{\alpha} + R_2^{\alpha} + \ldots + R_n^{\alpha})$$

The assumption of uniform rainfall within one mile either side of a rain gage is most surely inaccurate. However, it permits an approximate solution to the problem of attenuation due to rainfall that is not inconsistent with the measured values. Certainly a better correlation would have been obtained if the rain gages had been spaced closer together.

A number of rain events were analyzed and the data reduced to the equivalent 2-mile rate assuming uniform rainfall over the 2-mile spans. Fig. 11 is a scatter diagram showing transmission loss in decibels per mile due to precipitation versus precipitation in inches per hour. For comparison, Rydes' equation is plotted using the constant values suggested earlier.

The recording equipment at Mobile and Axis was arranged to record receiver input levels from 0 to 40 db below the normal input level [approximately −32 dbm (decibels referred to 1 milliwatt)]. The received signal strength was 40 db or more below the normal received level due to rainfall for 0.106% of the year on the Mobile–Mount Vernon path and 0.020% of the year on the Axis–Mount Vernon path. These figures indicate the expected order of outage time due to rainfall for single-hop 11-kmc radio systems having a 40-db fading margin and operating over similar paths in the Gulf Coast region. Fading margin is taken to mean the number of decibels the receiver input level might be reduced without the noise exceeding the system objective; outage time is defined as the time the noise does exceed the objective. Any predictions based on these figures for outage time would be pessimistic for most other areas of the United States since they reflect the attenuations due to severe rainfall found along the Gulf.

Implications of System Engineering

Rain attenuation is obviously a large factor in determining system reliability, and hence it reacts strongly on both the design and the application of the system. Since rainfall varies greatly in frequency and intensity from one region to another, it is important to be able to predict performance in any region so that the system as designed will have the widest possible application consistent with cost, and so that the applications engineer will know how to tailor those system parameters at his command to produce the degree of reliability desired.

It is not feasible for reasons of cost to measure rain attenuation in all parts of the United States, so it is necessary to use what rainfall data are available, and to couple the data, through what are thought to be reasonable assumptions, to

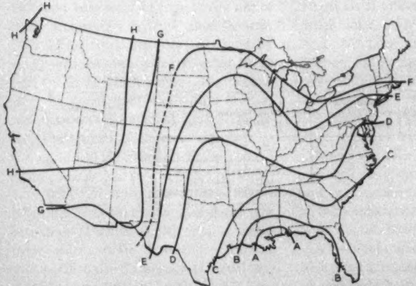


Fig. 12. Contours of constant path length for fixed outage time

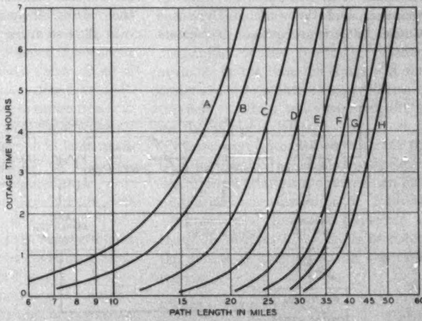
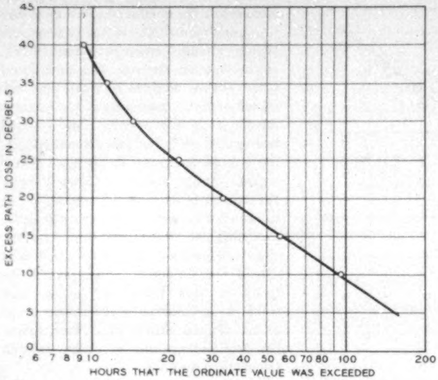


Fig. 13. Expected outage time in hours per year versus path length in miles for various areas of the United States



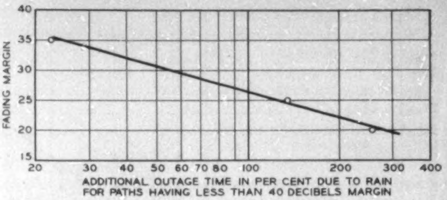


Fig. 14 (left). Excess path loss due to rainfall versus hours per year (at Mobile)

Fig. 15 (above). Additional outage time expected for 11-kmc systems having a fading margin less than 40 db

the relationships between rainfall and attenuation. The validity of the predictions rests clearly on the validity of the assumptions, and it is to be expected that further refinements in predicting rainfall outage will result from observing the performance of early 11-kmc systems.

As an approach to the problem of predicting outage time due to rainfall for all areas of the country, it has been assumed that the annual distribution of one-hour point rates is indicative of the annual distribution of instantaneous 30-mile path rates, along the lines suggested by Bussey.[5] This is equivalent to assuming a fixed storm pattern moving at 30 miles per hour in the direction of the path. Furthermore, it has been assumed that the frequency of occurrence of severe rainfall of the type measured in the Mobile area will be reduced in other parts of the country in proportion to the distribution of annual point rates of one inch or more per hour. Fig. 12, based on these assumptions and the work of Dych and Mattice,[9] illustrates contours of constant path lengths for fixed outage times for different areas of the United States. Fig. 13 shows the expected outage time due to rainfall for various path lengths in different rain areas of the United States. Curves *A* through *H* of Fig. 13 correspond to the general areas described by the contours on Fig. 12. The longer paths have been weighted somewhat to take account of less severe rainfall covering larger areas than storms typical of the Gulf region.

In engineering a complete 11-kmc radio-relay system, the rain outages of the individual hops must be added to obtain the performance for the system. Also, it is desirable to lay the system out in such a manner that the individual hops meet the same objective. From a practical standpoint this will not always be possible. Sometimes it is necessary to have one or more hops of a system electrically long; they will have insufficient fading margin and hence contribute more than their share of outage time. From the over-all system viewpoint this "excess" must be made up by imposing tighter requirements on the remaining hops.

To meet the over-all system objective it becomes necessary to know the contributions of the long hops, those having a fading margin less than 40 db. Fig. 14 shows excess path loss due to rain versus hours per year for the Mobile–Mount Vernon path. The shape of this curve is nearly identical with Bussey's curve on cumulative distribution for point rates at Washington, D. C. If the shape of this curve is assumed to be representative for other areas of the country, then the additional outage time for path lengths given by Fig. 13 can be estimated for hops having a fading margin less than 40 db. The data shown in Figure 14 have been rationalized and shown in Fig. 15 as an estimate of additional outage time.

Sometimes it is practical to shorten a proposed path to bring the fading margin up to 40 db. An approximation of the necessary reduction in path length can be made if uniform rainfall rate is assumed over the path. Under this condition Ryde shows the attenuation due to rainfall to be directly proportional to the path length. Thus the path length given

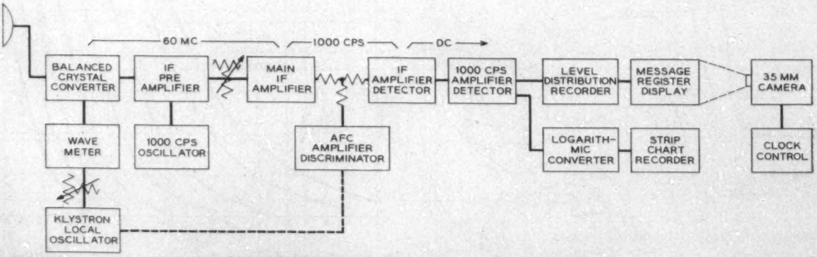


Fig. 16. Block schematic of radio receiver and recording equipment

in Fig. 13 can be shortened to correct for insufficient fading margin.

Rainfall in the extreme southeastern region of the United States will limit 11-kmc radio systems having a 40-db fading margin to path lengths of approximately 10 to 15 miles, depending on the number of hops if normal reliability objectives are to be met. Path lengths of 20 to 30 miles should be acceptable in the central area and path as long as 35 miles should be acceptable in the northwestern part of the country. However, in existing short-haul radio systems, the paths average 22 to 24 miles due to considerations other than those of propagation. It would then appear that 11-kmc systems will not be penalized unduly except in the extreme southeastern part of the United States.

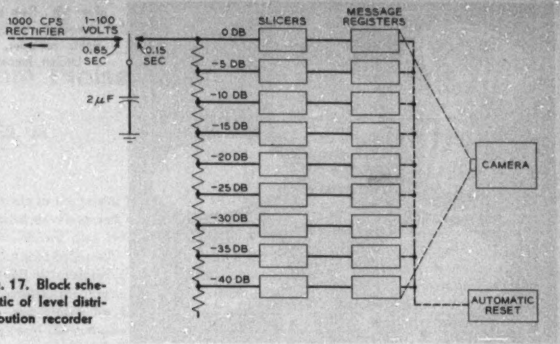


Fig. 17. Block schematic of level distribution recorder

Instrumentation

Transmitter

The transmitter employed a small commercially available klystron whose output was 0.5 watt. A variable probe was used to match the klystron to the waveguide to the antenna, and a 20-db directional coupler sampled the output so that frequency could be measured and the output power-monitored. The frequency stability was such that it was not necessary to use automatic frequency control. Since the a-c line was subject to frequent failure (a natural result of the thunderstorms whose rainfall provided the reason for the experiment), a strip-chart recorder with a mechanical clock drive was used to monitor the transmitter output.

Receiver

The receiver was adapted from equipment designed to record path loss at 4 kmc, which was in turn adapted from equipment designed to measure path reflections.[8] It is shown in block schematic in Fig. 16. (An extensive path-loss measuring program was carried out in 1947–1950 at 4 kmc prior to commercial use of those frequencies by the Bell System. This 4-kmc equipment was designed by H. C. Franke, and was converted to 11 kmc by S. D. Hathaway.) The normal received signal level was −32.2 dbm at Mobile and −31.1 dbm at Axis. The receiver was arranged to record signals from 0 db to −40 db relative to the normal signal.

A balanced converter supplied by a local klystron oscillator modulated the incoming 11.4-kmc signal to a 60-megacycle intermediate frequency. In the preamplifier the 60-megacycle signal was amplitude-modulated with 1,000 cps (cycles per second). The output of the preamplifier was divided between an intermediate-frequency amplifier feeding a frequency discriminator which provided automatic frequency control to the local klystron oscillator, and an amplitude detector where a 1,000-cps signal reasonably proportional to the microwave input signal over a 50-db range was recovered. The 1,000-cps signal was amplified and rectified at a level suitable to operate the display equipment.

Display Equipment

Two types of display equipment were used:

1. A level distribution recorder with a range of 40 db, which operated message registers which were photographed automatically on 35-millimeter film every 0.1 hour.
2. A strip-chart recorder with a logarithmic converter to produce a scale linear in decibels over a 50-db range.

Level Distribution Recorder

The electronic level distribution recorder to be described was developed in 1946 to replace a relay device that had been used for many years to study distributions of talker volume and telephone circuit noise. This circuit was conceived by L. Y. Lacy and developed by C.R. Eckberg for mobile use in connection with investigations of very-high-frequency transmission to vehicles, and later modified by H. C. Franke for use in microwave propagation measurements.

A series of nine d-c slicer circuits, each arranged to operate at an input of one volt, was connected to a voltage divider at 5-db intervals, thus covering a range of 40 db, as shown in Fig. 17.

A 2-microfarad capacitor was connected by a relay actuated by a sychronous timer to a cathode follower output of the 1,000-cps rectifier for 0.85 second. Then for 0.15 second the capacitor was connected to the voltage divider slicer circuit. The slicer circuits were arranged to drive message registers (counters) with the polarity such that if the slicer threshold was exceeded, the message register would not count; if the slicer threshold was not exceeded, the message register counted. Thus an input of 100 volts (normal signal) caused no counts. An input of 0.9 volt (41 db fade) caused all message registers to count. (Other level distribution recorders have been built with cross-gating between the slicers so that only the message register corresponding to the level just exceeded will operate. This type yields a histogram data presentation, whereas radio fading data are usually presented as a cumulative distribution.)

Standard 14-type telephone message registers were used, and a neon tube was connected across each message register for easy observation of individual counts. Some message register units were arranged with an automatic reset mechanism that reduced all message register readings to zero hourly. This simplified reducing the data since smaller numbers had to be dealt with, but the delicate mechanism of the resettable registers available at the time the equipment was designed led to maintenance problems in the field, so these units were rebuilt to use the simpler registers.

Means were provided to calibrate the level distribution recorder in a preliminary way from an accurate d-c source, but final calibration was always made from an accurate signal generator connected to the radio receiver input, so as to reduce the effects of nonlinearity in the radio receiver.

Camera Equipment

Pictures of the message register were taken every 0.1 hour by a 35-millimeter

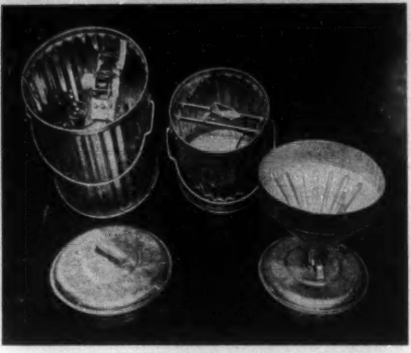


Fig. 18. Rain gage assembly, left to right: recorder, tilt bucket, funnel

camera adapted from a movie camera, under the control of a 24-hour electric clock which was included in the message register panel so that the time of each picture was recorded. Auxiliary message registers recorded the total number of measurements and the date, the latter register being actuated by the 24-hour clock. The message register units and cameras were enclosed in a trunk to reduce spurious illumination. A projection film reader was used to transfer the data to a data book for analysis.

LOGARITHMIC CONVERTER

The logarithmic converter accepted a d-c input voltage in the range −0.3 to −100 volts (as developed by the 1,000-cps detector) and changed it to a direct current proportional to the logarithm of the input voltage, with a range of 0–1 milliampere for the operation of a strip-chart recorder.

The d-c input voltage was chopped at a 60-cps rate and applied to a differentiating circuit followed by a d-c slicer circuit. The time that the differentiated wave exceeded the threshold of the slicer was proportional to the logarithm of the input voltage, so that the output of the slicer was a 60-cps wave with pulse-length modulation proportional to the input voltage in decibels. This output was filtered and applied to a strip-chart recorder to provide a linear 0–50-db recording. (It was found that imperfect filtering was desirable, in that a small amount of the 60-cps component improved the response of the strip-chart recorder to small changes in input.)

This display was limited by the slow response of the strip-chart recorder, whose time constant was about 0.5 second, so that the strip charts were used chiefly for monitoring and quick scanning of data.

RAIN GAGES

Automatic recording tilt-bucket rain gages, designed by L. E. Hunt, were placed approximately every 2 miles along the radio path, as shown in Fig. 2. The exact locations were determined by considerations of accessibility, since many of the roads across the radio path were little more than swamp traces. Also, an effort was made to minimize the effect of nearby objects such as trees and buildings.

The rain gage mechanisms, shown in Fig. 18, were proportioned so that the bucket tilted after each 0.01 inch of rain fell. A magnet attached to the tilt bucket closed the contacts of a glass-sealed switch during the instant of tilting. (Earlier models used open contacts which worked well in the laboratory but proved unsatisfactory under field conditions.)

RAIN GAGE RECORDERS

Because of the sparse wire facilities along the path, it was impossible to bring each rain gage circuit to a central recorder so individual strip-chart recorders were used at each rain gage. These used teletypewriter tape driven by a clock at the speed of 0.1 inch per minute (a little over 11 feet per day). Before this method was decided upon, the usual troubles of changes of paper dimensions caused much experimentation with chart drive; pins in the drive drums were superseded by a neoprene-band friction drive. Also, several types of clocks were tried: automobile electric clocks, precision automobile electric clocks, large spring-wound clocks, and finally governor-controlled d-c motors. None yielded the precision of timing desired, so that it was necessary to interpolate to avoid errors in correlating the individual rain gage records for particular rain events. (The improvements in the recorders were made by K. J. Frolund.) The strip-chart recorders also incorporated a stylus driven by an electromagnet that punched a tiny hole in the tape each time the rain gage bucket tilted. Thus, at the cloudburst rate of fall of 8 inches per hour, the punch marks were just over 0.01 inch apart, so that careful examination under a strong glass was required.

References

1. ATTENUATION OF CENTIMETER WAVES BY RAIN, HAIL, FOG AND CLOUDS, **J. W. Ryde, D. Ryde.** General Electric Company, Ltd., Wembley, England, May 1945.

2. THE ATTENUATION AND RADAR ECHOES PRODUCED AT CENTIMETER WAVE-LENGTHS BY VARIOUS METEOROLOGICAL PHENOMENA, **J. W. Ryde.** "Meteorological Factors in Radio Wave Propagation," The Physical Society, London, England, 1946, pp. 169–88.

3. THE RELATIONSHIP OF RAINDROP SIZE TO INTENSITY, **J. Otis Laws, Donald A. Parsons.** *Transactions*, American Geophysical Union, Washington, D. C., 1943, pp. 452–60.

4. STUDIES OF RAINDROPS AND RAINDROP PHENOMENA, **Wilson A. Bentley.** *Monthly Weather Review*, Washington, D. C., vol. 32, 1904, pp. 450–56.

5. MICROWAVE ATTENUATION STATISTICS ESTIMATED FROM RAINFALL AND WATER VAPOR STATISTICS, **Howard E. Bussey.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., July 1950, pp. 781–85.

6. PRECIPITATION ON THE MUSKINGUM RIVER WATERSHED BY 30-MINUTE INTERVALS. Soil Conservation Service, U. S. Department of Agriculture, Washington, D. C., 1938.

7. PORTABLE MICROWAVE TOWER. *Bell Laboratories Record*, New York, N. Y., vol. 2, no. 1, Jan. 1948, pp. 6–8.

8. PATH TESTING FOR MICROWAVE-RADIO ROUTES, **R. D. Campbell.** *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 326–34.

9. A STUDY OF EXCESSIVE RAINFALL, **H. D. Dych, W. A. Mattice.** *Monthly Weather Review*, vol. 69, Oct. 1941, pp. 293–301.

(See Discussion on Page 1061)

Performance of Class B Audio Amplifiers with Random Noise Signals

THERON USHER, JR.
ASSOCIATE MEMBER AIEE

DURING the past 3 years new problems in audio amplifier design have arisen as a result of using low-distortion high-power audio amplifiers for driving electrodynamic vibration exciters.[1] Modern vibration testing procedures require that random noise forces be used to excite the test specimens. This requirement has arisen from efforts to reproduce the actual environments occurring in airborne guided missiles.[2,3]

In order to obtain random vibratory forces from the vibration exciter, the currents and voltages produced by the power amplifier must also be random. Previous analyses of power amplifier performance have dealt only with the use of sinusoidal signals.[4,5] Due to the fundamental differences between random signals and sinusoidal signals,[6] the necessity for analyzing the performance of audio amplifiers for noise excitation becomes evident.

In particular, this paper considers the operation of the power stage of an audio amplifier which is excited by a band-limited random voltage. The power stage is assumed to contain triodes in class *B* push-pull operation. For simplicity of the driver circuit, the operation of the power stage is limited to the negative grid region. The power stage of the amplifier is coupled to the vibration exciter by means of a high-quality audio output transformer. For the purpose of keeping the results as general as possible, the vibration exciter is treated as a load with a complex impedance, the magnitude and angle of which vary widely over the frequency band of the noise signal.

The operational limitations of the amplifier power stage are discussed in the following sections, and are related mathematically to the noise voltages and currents in the power stage. The mathematical development draws heavily from noise theory and probability theory as introduced by Rice.[6] Finally, a graphical method is developed for selecting the output transformer turns ratio that provides the best match for the amplifier power stage and the complex impedance load.

Circuit Diagram and Nomenclature

Fig. 1 shows the circuit diagram for the power stage of the audio amplifier. Each triode is assumed to be biased to plate current cutoff, and restricted in operation to the negative grid region.

The various currents and voltages are defined in the following. In general, a lower-case symbol for voltage or current indicates an instantaneous value or a function of time. A bar above a time-varying quantity indicates the average or mean value of the quantity.

i_1 = plate current of upper triode, amperes (amp)
i_2 = plate current of lower triode, amp
v_{p1} = plate voltage of upper triode, volts
v_{p2} = plate voltage of lower triode, volts
v_1 = voltage across upper half of transformer primary, volts
v_2 = voltage across lower half of transformer primary, volts
i_b = plate supply current, amp
E_{bb} = plate supply voltage, volts
E'_{bb} = apparent plate supply voltage, volts
v = voltage across load impedance, volts
i = load current, amp
$Z(f)$ = complex load impedance $= R(f) + jX(f)$, ohms
N = transformer turns ratio, half primary to secondary
N_O = optimum transformer turns ratio
W_p = average plate power dissipation for both triodes, watts
σ_e = rms value of noise voltage at load, volts
σ_i = rms value of noise current at load, amp
R_p = plate resistance of power triode for zero grid bias, ohms
R_n = equivalent noise resistance of load, ohms
Z_n = equivalent noise impedance of load, ohms
$V(f)$ = spectral density of noise voltage at load, volts^2/cps (per cycles per second)
$I(f)$ = spectral density of noise current at load, amp^2/cps
$W(f)$ = spectral density for actual noise power at load, watt/cps
k = nondimensional level factor for spectra
$V'(f)$ = portion of $V(f)$ containing frequency variation, volts^2/cps
$I'(f)$ = portion of $I(f)$ containing frequency variation, amp^2/cps
P_c = probability of clipping or positive grid operation in power stage
K_1 = nondimensional parameter
K_2 = nondimensional parameter
y = nondimensional parameter

Limitations on Amplifier Operation

Two basic limitations on the operation of the power stage of the amplifier should be considered. First, the average plate dissipation of each power triode must not exceed the manufacturer's rating. The noise currents and noise voltages in the power stage may be assumed to be very nearly gaussian, and a mathematical expression for the plate dissipation can be derived.

The second limitation, that of negative grid operation, has previously been mentioned. This limitation will now be discussed in some detail. It is assumed that the signal voltages in previous stages of the audio amplifier are ideal, or have gaussian probability density functions, and that negligible clipping occurs in these stages. However, if the signal voltage has a large enough amplitude at the input to the power stage, either grid may draw current and cause excessive grid clipping in the power stage. Theoretically some clipping of the noise voltage and current always occurs if the input is perfectly gaussian, no matter how small the signal level, but clipping rapidly becomes insignificant as signal amplitude is reduced.

It can be seen that the objectives of high output and low distortion are not mutually compatible. A compromise between the two objectives may be achieved by having a specification that limits the clipping to a practical, but small, amount. The most convenient way to state this specification is to limit the probability of clipping, or of attempted positive grid operation, to a small value.

The probability approach to the problem of evaluating the amount of clipping

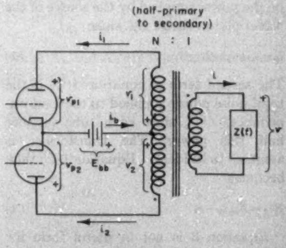


Fig. 1. Circuit diagram for the amplifier power stage

Paper 58-1237, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 4, 1958; made available for printing August 26, 1958.

THERON USHER, JR., is with Yale University, New Haven, Conn.

The work described in this paper and in the thesis listed as reference 1 is an outgrowth of development work for high-power audio amplifiers done by The MB Manufacturing Company, New Haven, Conn. The author wishes to thank them for making publication of this paper possible.

is advantageous because of the complexity of the problem. Due to the random nature of the signal, and the variation of load impedance with frequency, the path of operation for voltage and current on the plate characteristics of the power stage is nonrepetitive. This consideration alone suggests probability theory as the only reasonable approach for providing an analytical solution to the problem.

In the following sections the two limitations for amplifier operation are analyzed mathematically. From the previous arguments, it can be seen that the power stage is operating in an optimum fashion when the conditions of rated average plate dissipation and rated probability of clipping occur together. A graphical method for determining the optimum transformer turns ratio, N_O, for the foregoing condition is also described.

Average Plate Power Dissipation

The average plate power dissipation of an electron tube has been shown by many electronics texts[7] to be the average of the product of the plate voltage as a function of time and the plate current as a function of time. Thus the average plate dissipation for both tubes in Fig. 1 is

$$W_p = \overline{v_{p1}i_1} + \overline{v_{p2}i_2} \tag{1A}$$

By symmetry,

$$\overline{v_{p1}i_1} = \overline{v_{p2}i_2}$$

and

$$W_p = \overline{2v_{p1}i_1} \tag{1B}$$

By reference to Fig. 1, and basic theorems of probability theory concerning mean or average values,[8] the following simplifications of equation 1(B) may be accomplished.

$$W_p = 2\overline{(E_{bb}-v_1)i_1} = 2E_{bb}\bar{i}_1 - 2\overline{v_1 i_1} \tag{1C}$$

The first term of equation 1(C) is seen to be the power supplied by the source of the direct plate voltage E_{bb}, since

$$\bar{i}_b = \bar{i}_1 + \bar{i}_2 = 2\bar{i}_1 \tag{2}$$

The second term of equation 1(C) is the total noise power supplied to the load impedance, $Z(f)$, since each tube supplies half this power. The transformer is assumed to be ideal. Equation 1(C) then becomes

$$W_p = E_{bb}\bar{i}_b - \overline{vi} \tag{3}$$

Equation 3 is not in useful form for computational purposes. To obtain a useful form, the terms i_b and $\overline{vi}$ must be evaluated. The plate supply current can be expressed as a function of the load current:

$$i_b = \frac{1}{N}|i| \tag{4}$$

If the load current, i, is gaussian, or very nearly gaussian, the average plate supply current is defined by equation 5(A)

$$\bar{i}_b = \int_{-\infty}^{\infty} \frac{1}{N}|i| \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left[-\frac{i^2}{2\sigma_l^2}\right] di \tag{5A}$$

The integral in equation 5(A) has been evaluated,[8] and the average plate supply current is

$$\bar{i}_b = \sqrt{\frac{2}{\pi}}\frac{\sigma_l}{N} \tag{5B}$$

The rms load current σ_l is defined by the integral in equation 6, where $I(f)$ is the spectral density of the load current.

$$\sigma_l = \left[\int_0^\infty I(f)df\right]^{1/2} \tag{6}$$

Since $\bar{i}_b$ has been evaluated in terms of σ_l, it would be convenient to evaluate $\overline{vi}$ in terms of σ_l also. The physical significance of $\overline{vi}$ is that it is the average noise power supplied to the load. Mathematically, $\overline{vi}$ is a second moment of the joint probability density function for v and i. However, the evaluation of $\overline{vi}$ will make use of its physical significance.

The load impedance, $Z(f)$, can be resolved into its resistive and reactive components, both of which may be functions of frequency:

$$Z(f) = R(f) + jX(f) \tag{7}$$

The power dissipated per cps is then given by equation 8

$$W(f) = I(f)R(f) \tag{8}$$

The total noise power delivered to the load is the integral of equation 8 over the entire frequency band.

$$\overline{vi} = \int_0^\infty W(f)df = \int_0^\infty I(f)R(f)df \tag{9}$$

The computation in equation 9 can be simplified somewhat by recognizing the fact that the basic behavior of $I(f)$ as a function of frequency is usually specified by the requirements of a particular problem. It is desirable, however, to have the level of $I(f)$ as high as possible, under the operational limitations previously mentioned. Thus $I(f)$ can be split into a portion $I'(f)$, which describes the shape of the current spectrum as a function of frequency, and a factor k, which describes the level.

$$I(f) = kI'(f) \tag{10}$$

With this consideration in mind, it is possible to define an "equivalent noise resistance," R_n.

$$R_n = \frac{\overline{vi}}{\sigma_l^2} = \frac{\int_0^\infty I(f)R(f)df}{\int_0^\infty I(f)df} = \frac{\int_0^\infty I'(f)R(f)df}{\int_0^\infty I'(f)df} \tag{11}$$

It is easily seen from equation 11 that R_n is not dependent upon the absolute level of the current spectral density, and may be evaluated if the shape of the current spectral density is specified and the load impedance characteristics are known. In other words, R_n is not a function of σ_l.

Equation 3 is converted to a useful form by substituting the results found in equations 5(B) and 11.

$$W_p = \sqrt{\frac{2}{\pi}}\frac{E_{bb}}{N}\sigma_l - R_n\sigma_l^2 \tag{12}$$

Equation 12 is expressed in terms of quantities that can be easily computed or measured. For rated plate dissipation conditions, $W_p = W_{pr}$, and equation 12 may be rearranged to yield N in terms of σ_l.

$$N = \frac{\sqrt{\frac{2}{\pi}}E_{bb}\sigma_l}{W_{pr} + R_n\sigma_l^2} \tag{13}$$

Equation 13 will be used in a later section for describing the method of obtaining the optimum value for the turns ratio N_o.

Probability of Clipping

In order to evaluate the probability of clipping, it is necessary to define the var-

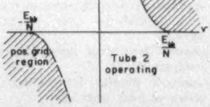


Fig. 2. Push-pull characteristics referred to secondary of output transformer

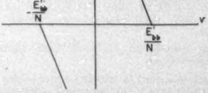


Fig. 3. Approximate zero grid voltage characteristics

ious regions of operation for the push-pull power stage. For analytical purposes, it is convenient to refer the push-pull characteristics to the load side of the output transformer, as shown in Fig. 2. The top half of the current-voltage plane corresponds to the condition when tube T_1 in Fig. 1 is operating and T_2 is cut off. The bottom half of the current-voltage plane corresponds to the opposite operating condition. The zero grid voltage curves in Fig. 2 are shown as dashed lines and the positive grid regions are shaded. The probability of clipping is evaluated by assuming ideal or gaussian probability density functions for load current and load voltage, and then computing the probability of finding both ideal variables jointly in the positive grid regions of Fig. 2.

Since v and i are related or nonindependent random variables, their ideal amplitude behavior is specified by the two-dimensional probability density function in equation 14.[8]

$$p(v,i)=\frac{1}{2\pi\sqrt{\sigma_e^2\sigma_i^2-\overline{vi}^2}}\times \exp\left[-\frac{\sigma_i^2v^2+\sigma_e^2i^2-2\overline{vi}\,vi}{2(\sigma_e^2\sigma_i^2-\overline{vi}^2)}\right] \quad (14)$$

It will be noted that the joint gaussian density function for v and i is completely described by the second moments of these variables, σ_e^2, σ_i^2, and $\overline{vi}$. The probability of finding both variables in the shaded areas of Fig. 2 is the double integral of the density function over these areas, as noted in equation 15.

$$P_c=\iint_{\substack{\text{shaded}\\ \text{areas}}} p(v,i)\,dv\,di \quad (15)$$

Due to the symmetrical nature of the problem, the integral in equation 15 can be simplified by integrating over the upper shaded area in Fig. 2, and doubling the results.

$$P_c=2\iint_{\substack{\text{upper}\\ \text{shaded}\\ \text{area}}} p(v,i)\,dv\,di \quad (15A)$$

The analytical problem is simplified greatly by approximating the zero-grid voltage curves in Fig. 2 by straight lines as shown in Fig. 3. The equation for the straight line in the upper half of the v-i plane is

$$v=\frac{E'_{bb}}{N}-\frac{R_p}{N^2}\,i \quad (16)$$

The voltage E'_{bb} is the "apparent plate supply voltage," and R_p is the plate resistance of the triodes. The straight-line approximation is very good except at low values of current. With equation 16, the limits of the double integral in equation 15(A) can be specified in a simple fashion. The differential segment of area in Fig. 3 further explains the specification of the limits. The result of combining equations 11, 14, and 15(A) is

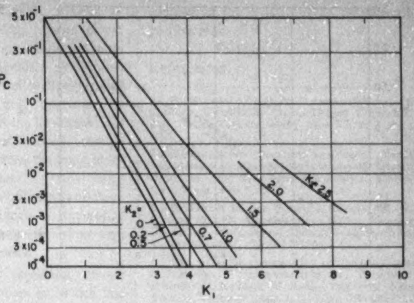


Fig. 4. Probability of clipping, P_c, versus K_1 for various values of K_2

$$P_c=2\int_0^\infty di\times \int_{\frac{E'_{bb}}{N}-\frac{R_p}{N^2}i}^{\infty}\frac{dv}{2\pi\sqrt{\sigma_e^2\sigma_i^2-R_n^2\sigma_i^4}}\times \exp\left[-\frac{\sigma_i^2v^2+\sigma_e^2i^2-2R_n\sigma_i^2vi}{2(\sigma_e^2\sigma_i^2-R_n^2\sigma_i^4)}\right] \quad (17)$$

The integrand in equation 17 is simplified by considering the definition for σ_e^2:

$$\sigma_e^2=\int_0^\infty V(f)df=\int_0^\infty I(f)|Z(f)|^2df \quad (18)$$

Next, the following definition for the "equivalent noise impedance" of the load is given:

$$Z_n=\frac{\sigma_e}{\sigma_i}=\left[\frac{\int_0^\infty V(f)df}{\int_0^\infty I(f)df}\right]^{1/2} =\left[\frac{\int_0^\infty kI'(f)|Z(f)|^2df}{\int_0^\infty kI'(f)df}\right]^{1/2} \quad (19)$$

It can be seen from equation 19 that Z_n is not dependent upon the magnitude of σ_e and σ_i, but is a constant, and may be calculated if $I'(f)$ and $Z(f)$ are known. With the result of equation 19, equation 17 becomes

$$P_c=2\int_0^\infty di\times \int_{\frac{E'_{bb}}{N}-\frac{R_p}{N^2}i}^{\infty}\frac{dv}{2\pi\sigma_i^2\sqrt{Z_n^2-R_n^2}}\times \exp\left[-\frac{v^2+Z_n^2i^2-2R_nvi}{2\sigma_i^2(Z_n^2-R_n^2)}\right] \quad (17A)$$

The integration process in equation 17(A) is further simplified by separating terms in the exponential and completing the square:

$$P_c=\frac{2}{2\pi\sigma_i^2\sqrt{Z_n^2-R_n^2}}\times \int_0^\infty \exp\left[-\frac{Z_n^2i^2-R_n^2i^2}{2\sigma_i^2(Z_n^2-R_n^2)}\right]di\times \int_{\frac{E'_{bb}}{N^2}-\frac{R_p}{N^2}i}^{\infty}\exp\left[-\frac{v^2-2R_nvi+R_n^2i^2}{2\sigma_i^2(Z_n^2-R_n^2)}\right]dv$$

$$=\frac{2}{2\pi\sigma_i^2\sqrt{Z_n^2-R_n^2}}\times \int_0^\infty \exp\left[-\frac{i^2}{2\sigma_i^2}\right]di\times \int_{\frac{E'_{bb}}{N}-\frac{R_p}{N^2}i}^{\infty}\exp\left[-\frac{(v-R_ni)^2}{2\sigma_i^2(Z_n^2-R_n^2)}\right]dv \quad (20)$$

Let

$$x=\frac{v-R_ni}{\sigma_i\sqrt{Z_n^2-R_n^2}}$$

and

$$y=\frac{i}{\sigma_i} \quad (21)$$

Then equation 20 becomes

$$P_c=2\int_0^\infty \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{y^2}{2}\right]dy\times \int_{K_1-K_2y}^\infty \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]dx \quad (22)$$

where K_1 and K_2 are nondimensional and are given by

$$K_1=\frac{E'_{bb}}{N\sigma_i\sqrt{Z_n^2-R_n^2}} \quad (23A)$$

and

$$K_2=\frac{\frac{R_p}{N^2}+R_n}{\sqrt{Z_n^2-R_n^2}} \quad (23B)$$

Related forms of the integral over x in equation 22 are tabulated,[9] but the defini-

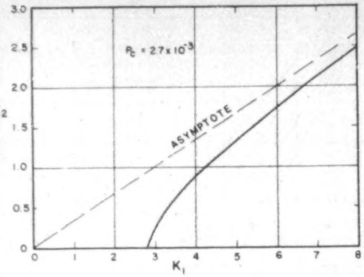


Fig. 5 (left). K_2 versus K_1 for constant P_c

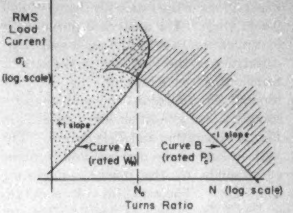


Fig. 6 (right). Graphical determination of optimum performance

tion for the normal probability distribution function will be used in equation 24(A).

$$F(\alpha)=\int_{-\infty}^{\alpha}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]dx \qquad (24A)$$

and

$$F(\alpha)=1-F(-\alpha) \qquad (24B)$$

In terms of the function $F(\alpha)$, which can easily be determined from tabulated integrals, the probability of clipping in equation 22 becomes

$$P_c(K_1,K_2)=2\int_0^{\infty}\frac{1}{\sqrt{2\pi}}\times \left\{\exp\left[-\frac{y^2}{2}\right]\right\}\{F[K_2y-K]_1\}dy \qquad (23)$$

It can be seen from equation 23 that P_c is a function of the nondimensional variables K_1 and K_2 only.

The integral in equation 23 has defied analytical evaluation for general values of K_1 and K_2, but has been evaluated numerically for several values of K_1 and K_2. The results are shown in Fig. 4, which shows P_c as a function of K_1, for various constant values of K_2. Two special cases of the integral in equation 23 have been evaluated. For $K_2=0$,

$$P_c(K_1,0)=F(-K_1) \qquad (24)$$

Also, when $Z_n\rightarrow R_n$ for a purely resistive load,

$K_1,\ K_2\rightarrow\infty$

but

$$\frac{K_1}{K_2}=\frac{E'_{bb}}{N\sigma_l\left(\frac{R_p}{N^2}+R_n\right)} \qquad (25)$$

and

$$\lim_{\substack{K_1\rightarrow\infty\\K_2\rightarrow\infty}} P_c(K_1,K_2)=2\left[1-F\left(\frac{K_1}{K_2}\right)\right]$$

$$=2F\left(-\frac{K_1}{K_2}\right) \qquad (26)$$

It has been mentioned that the probability of clipping must be limited to a small value in order that distortion be reasonably small. A typical value for P_c is 2.7×10^{-3}. This corresponds to the probability of a single gaussian random variable assuming a value greater than three times the rms value. The parameter K_2 is plotted as a function of K_1 for this typical value of P_c in Fig. 5. Equations 23(A) and 23(B) will be used with Fig. 5 in the next section to describe the graphical method for determining the optimum turns ratio, N_O.

Graphical Determination of Optimum Operating Conditions

As indicated previously, the power stage operates in an optimum fashion when the conditions of rated plate dissipation and of rated probability of clipping are satisfied simultaneously. The parameter used in the previous analysis which is indicative of the amplifier output is the rms load current, σ_l. When the amplifier is operating in optimum fashion, σ_l will have its largest possible value. The problem yet remains to determine the values of the amplifier parameters (E_{bb}; E'_{bb}; R_p; N) which will yield the maximum possible value of σ_l with regard to the specified amplifier limitations and the given load impedance.

Actually, a slightly more restricted aspect of this general problem will be considered. It will be assumed that the plate supply voltage and vacuum tube type have already been selected. This fixes the value of E_{bb}, E'_{bb}, W_{pr}, and R_p. The problem then becomes one of selecting the value of the transformer turns ratio that yields maximum output or maximum rms load current. Due to the complicated equations which describe the plate dissipation limitation and the signal clipping limitation, the solution can best be done graphically.

From equation 13, a graph of σ_l as a function of N may be constructed if E_{bb}, W_{pr}, and R_n are known. The shape of the typical curve resulting from the solution of equation 13 is shown on logarithmic co-ordinates in Fig. 6 and is labelled A. Points in the dotted area in Fig. 6 yield values for plate dissipation which are greater than the rated value. The shape of curve A can be deduced from equation 13.

At low values of σ_l, the $R_n\sigma_l^2$ term in equation 13 is negligible and N varies directly with σ_l. Thus the slope of curve A at low values of σ_l is very nearly unity. As σ_l increases, the slope increases and N reaches a maximum value. The region of negative slope for the rated plate dissipation curve is rarely of interest.

The next part of the graphical procedure involves finding and plotting the values of σ_l and N which satisfy the condition for rated probability of clipping. This process is more complicated than the previous one, and requires the combined use of equations 23(A), 23(B), and Fig. 5. The steps in the solution are as follows:

1. Values for N are assumed, and corresponding values for K_2 are computed from equation 23(B).
2. Values of K_1 corresponding to the values of K_2 in step 1 are found from Fig. 5.
3. Equation 23(A) is rearranged to yield

$$\sigma_l=\frac{E'_{bb}}{K_1N\sqrt{Z_n^2-R_n^2}} \qquad (27)$$

The known values of K_1 and N are then used to determine corresponding values of σ_l.

The graph of σ_l versus N may then be plotted. The shape of the typical curve resulting from the previous three steps is shown also in Fig. 6 and is labelled B. The points in the shaded area above curve B yield values for the probability of clipping which exceed the specified value. The shape of curve B in Fig. 6 can also be described by reference to equations 23(B) and 27 and also Fig. 5.

As N becomes very large K_2 approaches a fixed lower limit:

$$\lim_{N\to\infty} K_2 = \frac{R_n}{\sqrt{Z_n^2 - R_n^2}} \qquad (28)$$

For the specified probability of clipping, K_1 approaches a limit given by Fig. 5. From equation 27, σ_t and N are inversely related for high values of N, because K_1 approaches a constant value. This corresponds to a negative unity slope on logarithmic co-ordinates. As N decreases the magnitude of the slope of curve B in Fig. 6 decreases because K_1 increases in equation 27.

The intersection of curve A and curve B in Fig. 6 yields the maximum value of the rms load current obtainable from the amplifier without exceeding the rated plate dissipation of the power triodes or the specified probability of clipping the output signal. The corresponding value of N is the optimum turns ratio for the output transformer.

Conclusions

The evalution of the performance of push-pull triode power amplifiers in class B operation with random signals has been shown to be based on two reasonable limitations. The average plate power dissipation limitation is exactly the same as that for sinusoidal signals. The limitation involving the amount of clipping or of attempted positive grid operation is roughly comparable to the limitation imposed by the clipping of sinusoidal signals due to grid current. Thus the basic ideas presented in this paper are not really new. However, the application of these ideas for random signals is somewhat more complicated than the application for sinusoidal signals.

Although this paper specifically covers the analysis of amplifiers in which the operation is limited to the negative grid region, the results are not limited in this respect.

It is possible to analyze power amplifiers operating partially in the positive grid region, if the boundaries of the allowable region of operation can be approximated by straight lines as shown in Fig. 3 and given by equation 16. The amplifier limitation then becomes one of grid power dissipation or of available power from the driver stage instead of clipping of the signal due to the flow of grid current.

References

1. Performance of Class B Amplifiers with a Complex Impedance Load, **Theron Usher, Jr.** *D.E. Thesis*, Yale University, New Haven, Conn., 1957.

2. Establishing Vibration and Shock Tests for Missile Electronic Equipment, **M. Gertel.** *Report no. 230c*, The Barry Corporation, Watertown, Mass., Feb. 1, 1955.

3. Vibration in the Viking 9 Rocket, **M. W. Oleson, C. W. Cunningham.** *Rocket Research Report no. XVI*, Naval Research Laboratories, Washington, D. C., Nov. 19, 1954.

4. Graphical Determination of Performance of Push-Pull Audio Amplifiers, **B. J. Thompson.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 21, Apr. 1, 1933, pp. 591–600.

5. Triode Power Amplifiers with Complex Load Impedance, **W. J. Cunningham, J. Skalnik.** *Nonr 433(00), Report no. 1*, Office of Naval Research, Washington, D. C., Apr. 1952.

6. Mathematical Analysis of Random Noise, **S. O. Rice.** *Bell System Technical Journal*, New York, N. Y., vol. 14, Jan. 1945, pp. 46–160.

7. Theory and Application of Electron Tubes (book), **H. J. Reich.** McGraw-Hill Book Company, Inc., New York, N. Y., 1939.

8. Mathematical Methods of Statistics (book), **H. Cramér.** Princeton University Press, Princeton, N. J., 1946.

9. Tables of Normal Probability Functions. "Applied Mathematics Series," vol. 23, National Bureau of Standards, Washington, D. C., 1953.

A Stable Direct-Coupled Transistor Servo Preamplifier

ALBERT N. DeSAUTELS
ASSOCIATE MEMBER AIEE

Synopsis: In the designer approach to transistor servo-amplifier design, only limited consideration has been given to direct coupling. This is particularly true in applications involving operation over a wide temperature range. This paper discusses design and performance relationships and d-c stability considerations for a 2-stage direct-coupled preamplifier capable of d-c operating-point stability and high gain at temperatures in excess of 125 C (degrees centigrade).

In the conventional design of preamplifiers for servo applications, the engineer considers three types of coupling: transformer coupling, R/C coupling, and direct coupling.

In the comparison of the three types shown in Fig. 1, it is apparent that direct coupling has definite merit in miniaturization and economy. However, characteristics of interchanging transistors have generally led circuit designers to give only limited consideration to direct coupling.

Any variation in the d-c conditions of one stage of the conventional direct-coupled amplifier of Fig. 1 creates appreciable interaction with the other stage. This effect can be remedied to a certain extent by introducing considerable d-c degeneration in the emitter circuit. This is done by making the emitter circuit d-c resistance large compared to the collector circuit resistance. Thus, as the collector current increases with temperature increase, this current increase will show up more effectively as an increase in voltage potential at the emitter. With the base voltage fixed, an increase in potential on the emitter will reduce the emitter to base voltage and thus reduce the collector current. The result is a form of temperature compensation that performs fairly well.

However, the problem of interchangeability still exists. If all transistors had the same characteristics, they could be freely interchanged. Since this is not the case, to achieve a-c gain and d-c bias stability with interchanging transistors would require careful selection of transistors, additional stabilizing networks, etc., which, in many applications, is just not practical.

Limitations are also imposed by the use of the tantalum by-pass capacitors due to the temperature characteristics of the capacitors. At low temperatures a-c series resistance of the tantalum increases exponentially and presents a very significant degenerative effect.

A Simplified Direct-Coupled Circuit

A circuit which emphasizes to a high degree the desirable characteristics of small size, simplicity, and economy associated with direct coupling while still achieving excellent a-c stability is

---

Paper 58-1238, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 28, 1958; made available for printing September 18, 1958.

Albert N. DeSautels is with Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

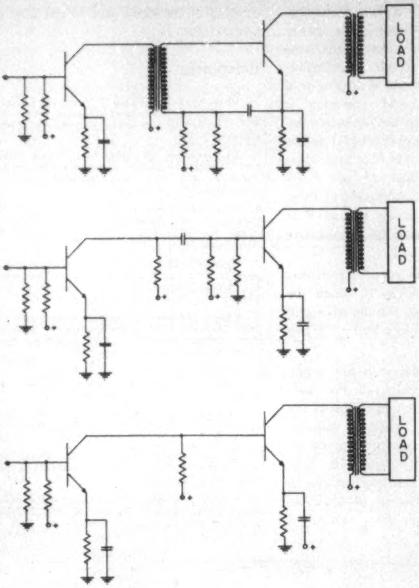


Fig. 1 (left). Conventional transistor preamplifiers

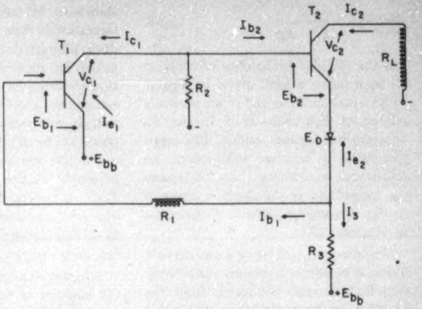


Fig. 3 (above right). Design analysis of direct-coupled preamplifier

shown in Fig. 2. This circuit uses a form of d-c feedback to achieve d-c stability.

Note the simplicity of the design. There are no tantalum capacitors and fewer components, yet this preamplifier is d-c stable with interchanging transistors and is capable of providing stable gain with silicon transistors to temperatures in excess of 125 C.

The theory of stabilization and bias is straightforward. The Zener diode, D_1, utilizes reverse diode characteristics to provide a constant emitter voltage for the second stage, T_2. This diode also determines the d-c collector voltage for the first stage, T_1. The d-c voltage drop across R_1 acts as a generator to provide bias current for T_1. The voltage drop across R_1 establishes the base bias voltage for T_2. When interchanging transistors with slightly different d-c characteristics or when collector currents increase with higher temperature, the circuit should automatically stabilize itself.

If I_{c_1} increases with temperature rise, the voltage drop across R_2 will rise. This means the base voltage of T_2 will decrease with respect to ground. Since emitter voltage of T_2 is constant (due to Zener diode), decrease in base voltage will tend to decrease T_2 collector current. Thus, the drop across R_1 will decrease, tending to reduce the base bias current for T_1 and thus reduce I_{c_1}. The same theory applies for increase in I_{c_2}.

This d-c stability is obtained by means of a tight d-c feedback loop around the two stages. This is provided by L_1 which has low d-c resistance but high a-c impedance. Therefore, there is tight d-c coupling, but practically no a-c feedback.

General Design Relationships

General design equations can be established which will relate component values to power source amplitude and d-c operating points.

From Fig. 3, the following relations are observed:

$$I_{b_1}+I_{c_1}+I_{e_1}=0 \tag{1}$$

$$I_{b_2}+I_{c_2}+I_{e_2}=0 \tag{2}$$

$$I_{b_1}+I_3+I_{e_2}=0 \tag{3}$$

$$I_{b_2}=\frac{I_{c_2}}{\beta_2},\ I_{b_1}=\frac{I_{c_1}}{\beta_1} \tag{4}$$

where β_1 and β_2 are common emitter current gains of T_1 and T_2, respectively.

Taking voltage loop sums and considering all voltages and currents as absolute values:

$$E_{bb}=V_{c_1}+R_2(I_{c_1}+I_{b_2}) \tag{5}$$

$$E_{bb}=I_3R_3+E_D+V_{c_2}+I_{c_2}R_L \tag{6}$$

$$E_{bb}=I_3R_3+E_D+E_{b_2}+R_2(I_{c_1}+I_{b_2}) \tag{7}$$

Solving equations 1 through 4 for I_3,

$$I_3=\frac{I_{c_2}}{\beta_2}+I_{c_2}-\frac{I_{c_1}}{\beta_1} \tag{8}$$

Inserting equation 8 in equation 6 and solving for R_3,

$$R_3=\frac{E_{bb}-E_D-V_{c_2}-I_{c_2}R_L}{I_{c_2}+\frac{I_{c_2}}{\beta_2}-\frac{I_{c_1}}{\beta_1}} \tag{9}$$

From Fig. 3,

$$I_3R_3=I_{b_1}R+E_{b_1} \tag{10}$$

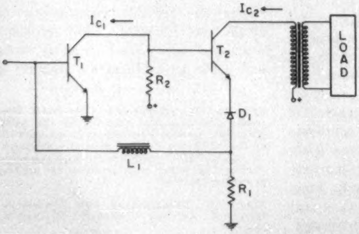


Fig. 2. The d-c stable direct-coupled amplifier

Vol,
Pt. 1
195

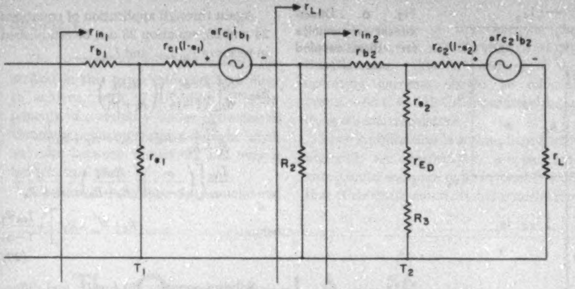


Fig. 4. Equivalent circuit for 2-stage direct-coupled preamplifier

Solving equation 10 for R_1 and submitting equation 8,

$$R_1=\frac{R_3\left(\frac{I_{c_2}}{\beta_2}+I_{c_2}-\frac{I_{c_1}}{\beta_1}\right)-E_{b_1}}{\frac{I_{c_1}}{\beta_1}} \tag{11}$$

From equations 5 and 4,

$$R_2=\frac{E_{bb}-V_{c_1}}{I_{c_1}+\frac{I_{c_2}}{\beta_2}} \tag{12}$$

From Fig. 3,

$$V_{c_1}=E_{b_2}+E_D+I_{b_1}R_1+E_{b_1} \tag{13}$$

Assuming $E_{b_2}=E_{b_1}$,

$$V_{c_1}=2E_{b_1}+E_D+I_{b_1}R_1 \tag{14}$$

Solving equation 14 for E_D and applying equation 4,

$$E_D=V_{c_1}-2E_{b_1}-\frac{I_{c_1}}{\beta_1}R_1 \tag{15}$$

For practical values of R_1, the right-hand term of equation 15 is negligible compared to the other voltages and equation 15 simplifies to

$$E_D\cong V_{c_1}-2E_{b_1} \tag{16}$$

DESIGN SUMMARY

Knowing V_{c_1}, I_{c_1}, V_{c_2}, I_{c_2}, E_{bb}, α_1, α_2, and R_L (d-c load resistance), component values for either *p-n-p* or *n-p-n* 2-stage direct-coupled amplifiers can be calculated from the following relationships using absolute values for current and voltage:

$$E_D=V_{c_1}-2E_{b_1} \tag{16}$$

$$R_3=\frac{E_{bb}-E_D-V_{c_2}-I_{c_2}R_L}{I_{c_2}+\frac{I_{c_2}}{\beta_2}-\frac{I_{c_1}}{\beta_1}} \tag{9}$$

$$R_1=\frac{R_3\left(I_{c_2}+\frac{I_{c_2}}{\beta_2}-\frac{I_{c_1}}{\beta_1}\right)-E_{b_1}}{\frac{I_{c_1}}{\beta_1}} \tag{11}$$

$$R_2=\frac{E_{bb}-V_{c_1}}{I_{c_1}+\frac{I_{c_2}}{\beta_2}} \tag{12}$$

Performance

In terms of a-c performance, the equivalent circuit would appear as shown in Fig. 4. Two-stage relationships for performance analysis can be established from this circuit in combined terms. However, the resulting equations are too complex for practical usage. It is more convenient to make a stage by stage analysis using established amplifier stage relationships.

Through use of

$$A_v=\frac{\alpha r_c r_L}{r_c[r_e+r_b(1-\alpha)]+r_L(r_e+r_b)} \tag{17}$$

and

$$A_i=\frac{\alpha}{1-\alpha+\frac{r_L}{r_c}} \tag{18}$$

for a known a-c load impedance, r_L, and using average small signal parameters associated with the selected operating points, the voltage and current gain of T_2 is calculated. Then r_{in_2} is calculated from

$$r_{in_2}=(r_{b_2}+r_{e_2}+r_{ED}+R_3)\left[\frac{r_{c_2}+r_L}{r_{c_2}(1-\alpha_2)+r_L}\right] \tag{19}$$

The second-stage input impedance, r_{in_2}, in parallel with R_2 provides the first stage load impedance, r_L. With r_{L_1} known equations 17 and 18 can again be used applying small signal parameters associated with T_1. The amplifier input impedance, r_{in_1}, would be obtained from

$$r_{in_1}=r_{b_1}+\frac{r_{e_1}}{1-\alpha_1} \tag{20}$$

Typical performance figures for this amplifier would be

Table I. Collector-Current Temperature Characteristics

Temperature, C	I_{c_1}, Milliamperes	I_{c_2}, Milliamperes
25	0.63	1.5
125	0.75	1.68
−55	0.57	1.38

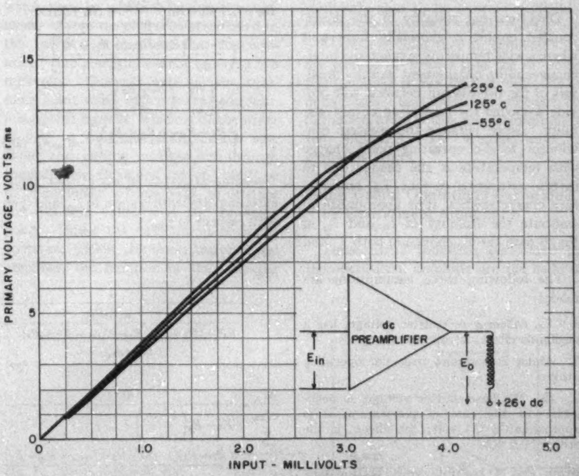


Fig. 5. Comparative response

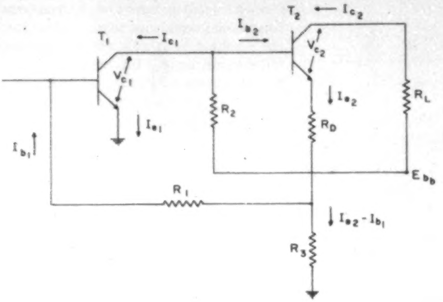


Fig. 6. Direct-current relationships for direct-coupled preamplifier

$A_v \cong 20{,}000$
$A_i \cong 500$
Power gain≅70 decibels
Input impedance≅2,500 ohms

TEMPERATURE PERFORMANCE

The graph of Fig. 5 presents comparative amplifier response at 25 C, 125 C, and −55 C as calculated for typical transistors. Stability of performance over the temperature range is indicated.

Table I presents comparative d-c collector current values as determined for the temperature extremes. Less than 15% variation in direct current from room temperature to temperature extremes is indicated.

Some D-C Stability Considerations

Direct-current stability of the direct-coupled circuit is especially important where a wide temperature range and interchangeability requirement exists. Analysis of the d-c biasing network will provide a means for measuring circuit stability. Since perhaps the greatest contributor to d-c operating point change with temperature is the change due to inverse saturation current, I_{co}, temperature characteristics, it is appropriate to associate the stability of I_{c_1} and I_{c_2} in Fig. 6 with the I_{co} change in both T_1 and T_2.

The following three assumptions are made:

1. V_c, collector to emitter voltage, has a negligible effect on collector current.
2. Alpha is constant over the operating range.
3. E_{be}, the base-emitter voltage, is negligible. In the case of germanium, it is approximately 0.1 volt; for silicon, in the order of 0.6 volt.

From Fig. 6, the following relationships are evident:

$$I_{e_1}=I_{b_1}+I_{c_1} \tag{21}$$

$$I_{e_2}=I_{b_2}+I_{c_2} \tag{22}$$

$$E_{bb}=(I_{c_1}+I_{b_1})R_2+I_{e_2}(R_D-R_3)-I_{b_1}R_3 \tag{23}$$

Since

$$\alpha I_e=I_c+I_{co}$$

or

$$I_e=\frac{I_c}{\alpha}-\frac{I_{co}}{\alpha} \tag{24}$$

then

$$I_b=I_c\frac{(1-\alpha)}{\alpha}-\frac{I_{co}}{\alpha} \tag{25}$$

Through application of equations 24 and 25, equation 23 can be established in terms of I_{c_1}, I_{co_1}, and I_{co_2},

$$E_{bb}=I_{c_1}\left[R_2-R_3\left(\frac{1-\alpha_1}{\alpha_1}\right)+(R_1+R_3)\times\left(\frac{1-\alpha_1}{\alpha_1}\right)\frac{R_2-R_2\alpha_2+R_D+R_3}{R_3}\right]+I_{co_1}\left[\frac{R_3}{\alpha_1}-\left(\frac{R_1+R_3}{\alpha_1}\right)\times\left(\frac{R_2-R_2\alpha_2+R_D+R_3}{R_3}\right)\right]-I_{co_2}R_2 \tag{26}$$

Solving equation 26 for I_{c_1} and differentiating, first with respect to I_{co_1} and then with respect to I_{co_2},

$$\frac{\partial I_{c_1}}{\partial I_{co_1}}=-\cfrac{1}{\cfrac{\alpha_1R_3R_2}{R_3^2-(R_1+R_3)(R_2-R_2\alpha_2+R_D+R_3)}-(1-\alpha_1)} \tag{27}$$

$$\frac{\partial I_{c_1}}{\partial I_{co_2}}=\cfrac{R_2}{R_2-R_3\left(\cfrac{1-\alpha_1}{\alpha_1}\right)+\cfrac{(R_1+R_3)(1-\alpha_1)(R_2-R_2\alpha_2+R_D+R_3)}{\alpha_1R_3}} \tag{28}$$

Again through application of equations 24 and 25, equation 23 can be established in terms of I_{c_2}, I_{co_1}, and I_{co_2},

$$E_{bb}=\frac{I_{c_2}}{\alpha_2}\left[\left(\frac{\alpha_1}{1-\alpha_1}\right)\left(\frac{R_3R_2}{R_1+R_3}\right)-\frac{R_3^2}{R_1+R_3}+R_2(1-\alpha_2)+R_D+R_3\right]-\frac{I_{co_2}}{\alpha_2}\left[\left(\frac{\alpha_1}{1-\alpha_1}\right)\left(\frac{R_3R_2}{R_1+R_3}\right)-\frac{R_3^2}{R_1+R_3}+(R_2+R_D+R_3)\right]+\frac{I_{co_1}R_3}{1-\alpha_1} \tag{29}$$

Solving equation 29 for I_{c_2} and differentiating, first with respect to I_{co_2} and then with respect to I_{co_1},

$$\frac{\partial I_{c_2}}{\partial I_{co_2}}=\frac{\left[\left(\frac{\alpha_1}{1-\alpha_1}\right)\left(\frac{R_3R_2}{R_1+R_3}\right)-\frac{R_3^2}{R_1+R_3}+(R_2+R_D+R_3)\right]}{\left[\left(\frac{\alpha_1}{1-\alpha_1}\right)\left(\frac{R_3R_2}{R_1+R_3}\right)-\frac{R_3^2}{R_1+R_3}+R_2(1-\alpha_2)+R_D+R_3\right]} \tag{30}$$

$$\frac{\partial I_{c_2}}{\partial I_{co_1}}=\frac{R_c}{\frac{1-\alpha_1}{\alpha_2}\left[\left(\frac{\alpha_1}{1-\alpha_1}\right)\left(\frac{R_3R_2}{R_1+R_3}\right)-\frac{R_3^2}{R_1+R_3}+R_2(1-\alpha_2)+R_D+R_3\right]} \tag{31}$$

From the four differential expressions, equations 27, 28, 30, and 31, the change in I_{c_1} can be expressed as a function of I_{co} change,

$$dI_{c_1}=\frac{\partial I_{c_1}}{\partial I_{co_1}}\,dI_{co_1}+\frac{\partial I_{c_1}}{\partial I_{co_2}}\,dI_{co_2} \tag{32}$$

and the change in I_{c_2} can be expressed in a similar fashion.

$$dI_{c_2}=\frac{\partial I_{c_2}}{\partial I_{co_1}}\,dI_{co_1}+\frac{\partial I_{c_2}}{\partial I_{co_2}}\,dI_{co_2} \tag{33}$$

If it is assumed that the I_{co} temperature characteristic is the same for both transistors, then $dI_{co_1}=dI_{co_2}$. The partial expressions of equations 32 and 33 may therefore be combined to produce stability factors relating the change in I_{c_1} and I_{c_2} to I_{co} change:

$$\frac{dI_{c_1}}{dI_{co}}=S_1=\frac{\partial I_{c_1}}{\partial I_{co_1}}+\frac{\partial I_{c_1}}{\partial I_{co_2}} \tag{34}$$

$$\frac{dI_{c_1}}{dI_{co}}=S_2=\frac{\partial I_{c_2}}{\partial I_{co_1}}+\frac{\partial I_{c_2}}{\partial I_{co_2}} \tag{35}$$

Ideally, S_1 and S_2 should approach zero. That is, the change in I_c due to one I_{co} will be cancelled by an opposing change in I_c due to the other I_{co}. Practically, in the amplifier discussed in this paper, S_1 is about 1 and S_2 is about 10.

Summary

The 2-stage transistor preamplifier described in this paper indicates an ability to achieve, with a minimum of components, d-c stability under extremes of transistor characteristics changes such as occur between transistors and over a temperature range.

Adequate d-c stability can be achieved under conditions of interchanging the transistors having a beta range of nearly three to one. Effective control of d-c operating currents should be effected over a −55 C to 125 C temperature range using silicon transistors.

Since stabilization is accomplished with negligible a-c degeneration, a-c performance should compare quite favorably with that of the more conventional transformer or *R/C* coupled preamplifiers. Design and stability factor relationships are applicable to *p-n-p* as well as *n-p-n* types of transistors.

Reference

1. A Comparison of Three Common Emitter Transistor Servo Preamplifiers, A. N. DeSautels. *AIEE Transactions*, vol. 75, pt. I, Mar. 1956, pp. 17–25.

The Operational Amplifier as a Laboratory Tool

PAUL E. PFEIFFER
ASSOCIATE MEMBER AIEE

THE "operational amplifier" as it is usually referred to in computer terminology has proved its versatility and usefulness as the backbone of modern electronic differential analyzers. In a rapidly increasing number of engineering offices, schools, and laboratories, these devices are available either in small computing installations or as separate plug-in units. It is the purpose of this paper to describe briefly their operation and to show how their usefulness may be extended beyond the usual computing applications. The applications described are suggested by a point of view which is certainly not new, but which does not seem to have been exploited to its fullest extent.

Characteristics and Function of Operational Amplifiers

The operational amplifier is normally used in a feedback arrangement as shown in Fig. 1. The amplifier itself is a high-gain d-c amplifier having a negative gain and a reasonably small output impedance, so that normal loads do not appreciably affect the voltage gain. The input to the amplifier at the junction of the input and feedback paths is known as the summing point. The voltage at this point will be referred to as the summing point voltage, and the current entering the amplifier will be referred to as the summing point current. These are designated by e_s and i_s in Fig. 1. It should be noted that all voltages referred to are potentials with respect to a fixed reference level taken as zero. This is often referred to as the signal ground.

Table I lists ideal characteristics of the operational amplifier and typical values which may be expected in actual equipment. From the characteristics listed in this table, it is apparent that the summing point potential cannot exceed a few millivolts. Currents entering the summing point must add to equal the summing point current, which is of the order of 0.01 μa (microamperes) to 100 $\mu\mu$a. Input and output voltages act through transmission elements, which may be quite nonlinear in some cases, to produce currents and voltages at the summing point. Under the assumption of stable operation within the linear range of the amplifier, the function of the amplifier may be idealized as follows. The output voltage adjusts automatically to maintain, to a high degree of approximation, the two conditions: 1. Zero net summing point current (current balance condition). 2. Zero summing point voltage.

This simplified view of the function of the operational amplifier aids the comprehension of known circuits and helps to visualize new possiblities. Once a circuit is conceived on the basis of the approximation, its operation may be studied with more precision by careful analysis which takes into account such factors as finite gain, drift, output impedance, stability considerations, etc. Often, however, it is sufficient to make simple approximate calculations based on the actual parameters to determine bounds on errors due to departure from the ideal. For example, if it is known that the summing point is less than 1,000 $\mu\mu$a, then the current balance must be within this figure. For many applications, this error is immediately seen to be entirely negligible. Or if it is known that output voltage will not exceed 20 volts, then an amplifier voltage gain of 10^4 insures that the summing point voltage is less than 2 millivolts in magnitude. This may be sufficient information to show that the assumption of zero summing point voltage leads to negligible errors.

Some Useful Operations Performed by Operational Amplifiers

Three useful functions performed by the operational amplifier are the basis of

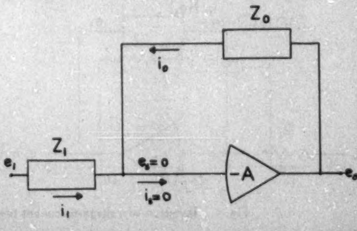


Fig. 1. Usual feedback configuration for operational amplifier

---

Paper 58-1198, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 16, 1958; made available for printing August 12, 1958.

Paul E. Pfeiffer is with The Rice Institute, Houston, Tex.

The author is grateful to a number of students who have tested various of the schemes described herein; especially to W. R. Peters, electronics technician in the Electrical Engineering Department, Rice Institute, for suggesting the resistance comparison scheme and putting into practice some of the ideas reported in this paper.

Table I. Operational Amplifier Characteristics

Items	Values		
	Ideal	Practical Range	
Voltage gain*	minus infinity	-10^4	-10^8
Input current	0	10^{-8}	10^{-11} amperes
Output impedance	0	300 ohms	100 ohms
Drift (referred to input)	0	2 millivolts	100 μvolts
Maximum load current		2 milliamperes	15 milliamperes
Maximum output voltage		±50 volts	±125 volts
Maximum summing point voltage	0	10 millivolts	1 millivolt

* With auxiliary chopper stabilizing amplifiers, d-c gain may be in excess of 10^7, giving summing point voltages at d-c of the order of 10^{-6} volts.

most of the techniques described in this paper:

1. Current measuring; see Fig. 2.
2. Current controlling; see Fig. 3.
3. Feedback function inversion; see Fig. 4.

These functions are, of course, simply ways of utilizing the inherent operation of the amplifier and feedback configuration, and are therefore closely related.

The current-measuring function is readily apparent from the requirements of the current balance condition, which requires that the current entering the summing point be equal to the current leaving by the feedback resistor path. This requires that for the circuit of Fig. 2 the relation $e_0=-i_1R_0$ must hold. Within the current-handling limits of the amplifier, this condition must be met regardless of load impedance. Hence a low-impedance voltmeter of a precision laboratory type may be used for current indication. It should be noted that the current is measured at a point essentially at signal ground potential. This has the advantage of producing small voltage drops in the current path. However, the current measured must be current to ground. If R_0 is sufficiently high, extremely small currents down to the order of 100 μμa may be indicated.

The measurement of extremely small currents, however, may be subject to a considerable number of errors due to drift voltages. Analysis shows this to be especially serious if the source impedance is small compared to the feedback resistor R_0. This situation, however, is not likely to occur, as an example will show. Suppose it is desired that a current of the order of 10^{-8} amperes be measured. A feedback resistor R_0 of 10 megohms would give a drop of 0.1 volt. This means that the source impedance would also have to be at least 10 megohms if the source voltage were as much at 0.1 volt. In the case of very low source voltages, much more detailed procedure and analysis are necessary to account for significant factors. In these cases, however, any measurement techniques are likely to be difficult.

The current-controlling function is obtained by making the input element a resistor R_1, and putting the device to be controlled in the feedback path, as in Fig. 3. Input current i_1 is very nearly proportional to input voltage e_1. Variations of the current i_1 from the value e_1/R_1 are due to the summing point voltage required to produce the necessary amplifier output voltage e_0. The magnitude of the i_1 variations from the desired amount are approximately e_0/AR_1. It is, of course, assumed that the system is stable. Generally this requires that an increase in the output voltage produces an increase in feedback current.

The feedback function inversion operation is a special case of the current control operation. Two cases are indicated in Fig. 4. In Fig. 4(A) it is assumed that the feedback element is such that its output current $i_0=f(e_0)$. If an inverse function exists, then

$$e_0=f^{-1}(i_0)=f^{-1}(-e_1/R_1) \tag{1}$$

If the feedback element is essentially a voltage transfer element, it may be converted to a current output element by

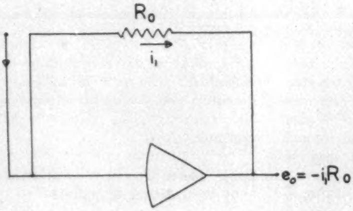


Fig. 2 (left). Operational amplifier as current-measuring device

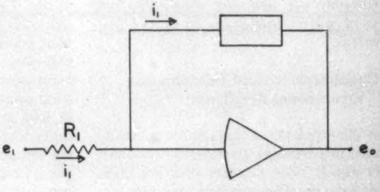


Fig. 3 (right). Operational amplifier as current-controlling device

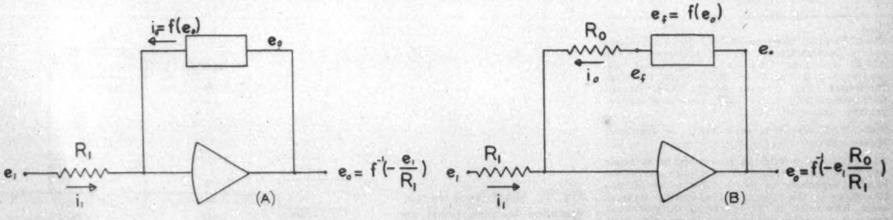


Fig. 4. Inversion of voltage-to-current transmission element and of voltage-to-voltage transmission element

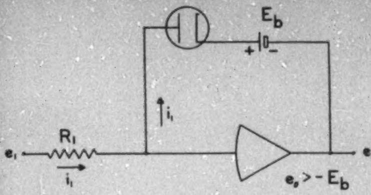


Fig. 5 (left). Logarithmic amplifier

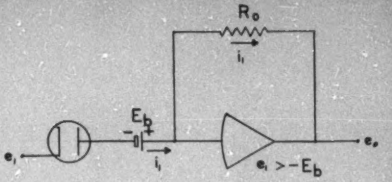


Fig. 6 (right). Exponential amplifier

adding a resistor R_0. Under the usual assumption of negligible summing point voltage, $i_0=e_f/R_0$. If $e_f=f(e_0)$, then

$$e_0=f^{-1}(-e_1R_0/R_1) \tag{2}$$

Applications

The following examples serve to illustrate the functions just described. Also it is believed that they are of interest in themselves. They have been employed in a number of applications in the laboratory.

LOGARITHMIC AMPLIFIER

The logarithmic amplifier diagrammed in Fig. 5 may be considered as an example of either current control or feedback function inversion.

If a thermionic vacuum diode is operated in the negative plate potential region, the current is an exponential function of the plate potential. This relationship may be quite precise over several decades of current range. If a biased diode is used as a feedback element, as shown in Fig. 5, the output voltage is proportional to the logarithm of the input voltage for several decades of the latter. The current balance equation for this circuit becomes

$$\begin{aligned} e_1/R_1 &= i_0 \exp\left[-\alpha(e_0+E_b)\right] \\ &= I_b \exp(-\alpha e_0) \end{aligned} \tag{3}$$

where $I_b=i_0\exp(-\alpha E_b)$. Equation 3 may be rewritten

$$\exp(-\alpha e_0)=e_1/(I_bR_1)=e_1/e_{1b} \tag{4}$$

Taking logarithms leads to the result

$$\begin{aligned} e_0 &= -(1/\alpha)\ln(e_1/e_{1b}) \\ &= -K_\alpha \log_{10}(e_1/e_{1b}) \end{aligned} \tag{5}$$

In equations 4 and 5, e_{1b} is the value of e_1 required to make e_0 equal to zero. The factor $K_\alpha=2.3026/\alpha$ is in volts per decade of current in the diode, or volts output per decade of input voltage change.

It is obvious that the logarithmic range cannot extend to zero since log x becomes negatively infinite as x goes to zero. Leakage resistance in the diode, amplifier saturation, etc., serve to provide the inevitable physical limits. It is also easily shown that the dynamic resistance of the diode becomes quite high for very small currents. Thus with small e_1 the amplifier operates with very small feedback and its sensitivity to stray pickup, internal noise, etc., becomes quite high. The gain in these regions may be limited at the expense of the range of logarithmic behavior by shunting the diode with a very high resistance.

The circuit has been used successfully with a pair of diodes in opposite directions to give logarithmic compression of a-c signals. It has also been used in a special-purpose computer involving several logarithmic and antilogarithmic amplifiers, with good over-all accuracy.[1]

AN EXPONENTIAL AMPLIFIER

If the biased diode utilized in the logarithmic amplifier is used as the input element rather than the feedback element, the resulting system provides an exponential relation between output and input voltage. The basic diode equation, Fig. 6, is

$$\begin{aligned} i_1 &= -i_0 \exp\left[-\alpha(e_1+E_b)\right] \\ &= -I_b \exp(-\alpha e_1) \end{aligned} \tag{6}$$

This results in output voltage e_0 given by

$$e_0=-i_1R_0=E_0\exp(-\alpha e_1),\ \text{for } e_1>-E_b \tag{7}$$

Here $E_0=I_bR_0$ is the output voltage when e_1 is zero.

By reversing the diode, the system may be used to handle positive voltages. The behavior is indicated by making appropriate sign changes in equations 6 and 7. Within the appropriate voltage ranges, good exponential behavior for two to three decades of output voltage is possible.

INVERSION OF VOLTAGE TRANSFER FUNCTIONS

The circuit of Fig. 7 shows a simple arrangement for inversion of a voltage transfer function. If $G(s)$ is the voltage transfer function $E_f(s)/E_0(s)$, the current balance condition leads directly to the result

$$\frac{E_0(s)}{E_1(s)}=\frac{-R_0}{R_1G(s)} \tag{8}$$

It should be noted that only minimum phase transfer functions are stable when inverted.

These techniques have been utilized in the study of certain unstable systems.[2] Unstable systems of known transfer functions are obtained by inverting nonminimum phase transfer functions. These in turn are stabilized by putting them in an appropriate feedback loop. The behavior of the unstable system is then studied by driving the outer loop and measuring within the loop.

MEASUREMENT OF VACUUM TUBE PARAMETERS

The use of vacuum tubes at low voltage and currents is receiving considerable attention.[3] Published data on tubes for such operation are limited and in many cases nonexistent. The measuring and controlling function of the amplifier plus its very small grid or summing point swing may be utilized to provide simple measurement techniques. Figs. 8 and 9 contain circuits for ready measurement of the basic parameters of a triode.

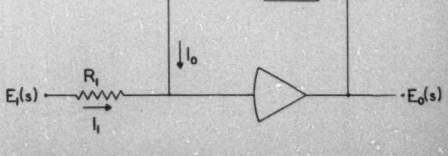


Fig. 7. Inversion of voltage transfer function

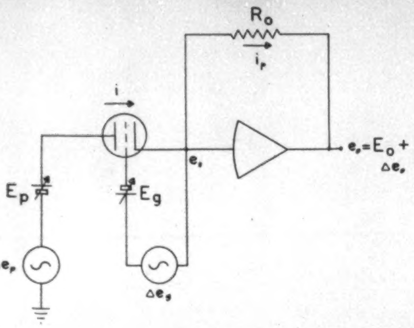


Fig. 8. Circuit for measuring transconductance and plate resistance

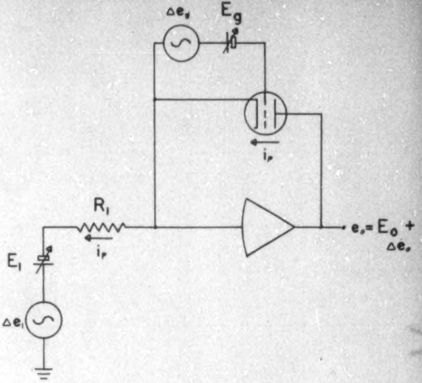


Fig. 9. Circuit for measuring amplification factor and plate resistance

Vol
Pt.
195

The circuit of Fig. 8 may be used to determine the transconductance g_m and the plate resistance r_p; the circuit of Fig. 9 yields measurements of amplification factor μ and the plate resistance r_p. Of course, any two of these parameters give the third by their well-known relationship $\mu = r_p g_m$.

Measurement of Transconductance

By definition

$$g_m = (\Delta i_p/\Delta e_g) \tag{9}$$

with e_p constant. In the circuit of Fig. 8 the plate voltage E_p and the grid bias E_g are fixed at the desired values. Measurement of the d-c component E_0 of the output voltage e_0 gives plate current I_p according to the relation $I_p = -(E_0/R_0)$.. Hence, static characteristics of the tube are readily obtained. An a-c component of grid voltage Δe_g is added in series with the bias voltage. Because of the negligible summing point voltage swing, the variation of E_p thus produced is negligible. Hence

$$g_m = -\Delta e_0/(\Delta e_g R_0) \tag{10}$$

By making $R_0 \Delta e_g$ some convenient power of 10, readings of Δe_0 are equal to $-g_m$ in corresponding units. For example, if R_0 is 0.1 megohm and Δe_g is 0.01 volt, then $-\Delta e_0$ is equal to g_m in thousands of micromhos. For convenience, Δe_g is made a low frequency sine wave so that the static components are read with d-c voltmeters and the changing portions are read on a-c instruments.

For many tubes, the circuit of Fig. 8 may be simplified by returning the bias voltage to ground rather than to the cathode. In this case, the voltage Δe_g differs from the applied voltage by the swing Δe_s of the summing point voltage. Often this is quite negligible. Also, the plate voltage differs from the applied voltage by Δe_s. This can nearly always be neglected. If the effect of Δe_s on plate voltage is neglected, the true value of g_m is obtained by multiplying the apparent values given in equation 10 by the factor $(1-E_A)$ where E_A is given by

$$E_A = 1/(1+A/R_0 g_m) \doteq R_0 g_m/A \tag{11}$$

Measurement of Amplification Factor

By definition

$$\mu = -(\Delta e_p/\Delta e_g) \tag{12}$$

with i_p constant. The circuit of Fig. 9 controls i_p and keeps it constant at the value $-E_1/R_1$. The potential Δe_1 is kept zero. The grid bias voltage E_g is set at the desired value and the d-c component E_0 of the output voltage e_0 is the plate voltage. Note that negative E_1 makes for positive E_p as required physically. Measurement of μ is a straightforward voltage ratio measurement as indicated in equation 12.

As in the circuit of Fig. 8, the grid voltage may be returned to the ground rather than to the summing point. Also, as in the previous case, the true value of μ is obtained by multiplying the apparent value by $(1-E_A)$ with E_A given in equation 11. This may be rewritten

$$E_A = \mu R_0/A r_p \tag{13}$$

Thus if R_0 is of the same order of magnitude as the plate resistance r_p, the error is approximately μ/A. For $A = 10^4$, a value of μ of 100 would be measured wth an error of about 1%, and in any case the error would tend to increase with μ.

Measurement of Plate Resistance

By definition

$$r_p = (\Delta e_p/\Delta i_p) \tag{14}$$

with e_g constant. Two possibilities pre-

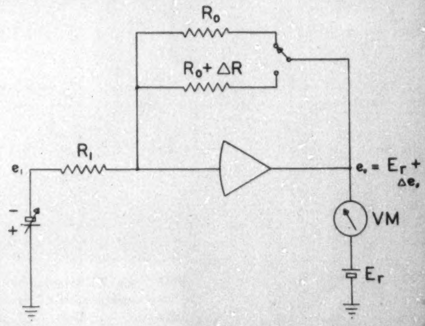


Fig. 10. Resistance comparison scheme

sent themselves. One is to make Δe_p the driving function and measure Δi_p. In the circuit of Fig. 8 this is accomplished in a straightforward manner. Thus

$$r_p = -(\Delta e_p/\Delta e_0)R_0 \tag{15}$$

Possibilities for direct reading by suitable choice of R_0 and Δe_0 are apparent. As in the case of the g_m measurements, static operating conditions are indicated.

If the circuit of Fig. 9 is used, Δi_p is made the driving function by controlling Δe_1 (with Δe_g held at zero). In this case

$$r_p = -(\Delta e_0/\Delta e_1)R_1 \tag{16}$$

For measurement of plate resistance, returning the grid voltage to ground will produce errors similar to those in the other parameters. Generally these errors are quite negligible for practical purposes and the simpler arrangement may be used satisfactorily.

It should be apparent that similar techniques can be used with transistors. This in fact has been done by one of the author's students.[4]

A D-C Resistance Comparison Circuit

A very simple resistance comparison scheme has been suggested by W. R. Peters. It works quite well for high resistances, up to possibly 100 megohms. For high resistances, oridinary bridges become quite insensitive, whereas this system maintains its null sensitivity. Also, in this system, the null indication is linear with respect to resistance variation. A simplified configuration is shown in Fig. 10. The operation is as follows: Suppose the resistance R_0 is taken as standard. Input voltage e_1 is adjusted until output voltage e_0 is equal to the reference or pedestal voltage E_r, as indicated by zero reading on the null voltmeter. Comparison resistor $R_0+\Delta R$ is then switched into the feedback path. The voltage equations in the two cases are respectively

$$E_r = -e_1(R_0/R_1) \tag{17}$$

and

$$E_r+\Delta e_0 = -e_1[(R_0+\Delta R)/R_1] \tag{18}$$

A little algebra shows that these relations imply

$$(\Delta e_0/E_r) = (\Delta R/R_0) \tag{19}$$

Notice that R_1 does not appear in the final expression. Also, the null voltage reading Δe_0 gives a direct linear measure of the variation of the comparison resistance from the standard. Care should be taken to protect a sensitive voltmeter during the switching instant, since momentarily the output is likely to vary considerably from E_r. This can be handled by proper switching, if desired.

Conclusions

The versatility of the operational amplifier is such that its usefulness is not limited to the usual computing applications, manifold and important as they are. The laboratory techniques described in this paper represent the more useful and interesting of the ideas which have been utilized by the author and his students. They were for the most part visualized in terms of the three basic operations of current measurement, current control, and feedback function inversion. Specific circuits, once conceived, may be analyzed more exactly by the usual techniques and methods of analysis. It is believed that the list of useful applications can be extended considerably by anyone who understands the basic operations. The principal limit on the applications is probably the current and voltage limitations of the operational amplifiers available.

References

1. An Application of Analog Computer Techniques to the Problem of Interpretation of the Contact Log, Arthur E. Nall, Jr. *M.S. Thesis*, The Rice Institute, Houston, Tex., 1957.

2. Unstable Linear Systems and the Minimum Phase Condition, Paul E. Pfeiffer. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 265, Apr. 1958, pp. 291–301.

3. Extension of Analog Computer Facilities for the Measurement of Vacuum Tube Characteristics, James S. Crabbe, Norman F. Rode. *Texas Engineering Experiment News*, College Station, Tex., vol. 8, Dec., 1957, pp. 10–12.

4. The Use of Operational Amplifiers in the Measurement of Transistor Parameters, William C. Hazel. *Review of Scientific Instruments*, New York, N. Y., vol. 29, Mar. 1958, pp. 235–37.

◆

Discussion

James R. Cox (A. and M. College of Texas, College Station, Tex.): Dr. Pfeiffer's organization of some of the less conventional uses of operational amplifiers should lead to new circuits employing these amplifiers. Several ideas similar to those described in the paper have been used at the A. and M. College of Texas where, however, the circuits have been treated on an individual basis, not by an organized method that would lead logically into new circuits.

The operational amplifier method of measuring tube parameters was found to be convenient and accurate. In the case of current measurements, chopper stabilization of the amplifier was found valuable. In some cases, a microammeter or galvanometer was used in the feedback loop. In this configuration, the amplifier effectively lowers the impedance of the meter to a few ohms. The current is read directly. Also, precision capacitors with low leakage, usually available in analog computers, can be substituted for the feedback resistor. This is convenient for millimicroampere currents when precision resistors of sufficient size are not available. The current is given by

$$i = C\frac{de_0}{dt} \tag{20}$$

or in the case of constant current

$$I = C\frac{\Delta E_0}{\Delta T} \tag{21}$$

The current is found by timing output voltage change with a stop watch.

A chopper stabilized amplifier and microammeter is being used as a voltage detector for a Wheatstone bridge. While its sensitivity compares with that of a galvanometer, it is more rugged, easier to use, and can be mounted on a radio rack panel. In addition, its sensitivity is continuously variable.

Dr. J. R. Biard has used the basic idea of the logarithmic and exponential amplifiers in an electronic multiplier. The multiplier is more versatile than other multipliers since it can raise variables to integral or nonintegral powers as well as multiply two variables. The operator, however, is burdened with a somewhat complicated alignment procedure.

T. N. Whitaker (University of Houston, Houston, Tex.): Dr. Pfeiffer views the operational amplifier as worthy of study in its own right; emphasizes its versatility and shows several noncomputing uses other than the usual one of analog computation. The great significance of this familiar use has tended to distract attention from finding the others. This paper should help restore balance.

It seems appropriate to expand somewhat on the less orthodox uses suggested by the author. Reference 2 of the paper explains the use of operational amplifiers in the experimental study of certain unstable systems. The major emphasis of this earlier paper was on the minimum phase condition, and only incidentally on the use of operational amplifiers. This is a rich vein to work, nonetheless, for it illustrates several different uses of amplifiers as laboratory tools.

It was desired to synthesize a nonminimum-phase system for laboratory demonstration of its unstable nature. Operational amplifiers were used at several steps in this synthesis because of their particular input-output characteristics. Since the non-minimum-phase transfer function may be represented as the tandem sum of a minimum-phase, plus an all-pass transfer function, the synthesis of an all-pass section was deemed necessary. This was accomplished, in part, through the use of a resistance-capacitance phase-shifting network. Instead of using a center-tapped transformer to provide the necessary 180-degree phase difference in the voltages applied to the two ends, an operational amplifier was used to invert the phase at one terminal. This combination simulates a near-perfect center-tapped transformer which has the remarkable characteristic of constant transmission down to zero frequency. Also, this is an application of the inversion of the simplest voltage transfer function, a constant.

Once the phase-lag circuit was synthesized, its transfer function was inverted by use of another amplifier, specifically mentioned in this paper. The resulting system was the system of interest in the study, but it was unstable and, in that form, not amenable to study. It was recognized that this system could be made a part of a larger, stable system by closing an appropriate feedback loop. A third operational amplifier made this possible. This was another noncomputing use of the amplifier, and an example of the fruitfulness of a point of view. The other applications mentioned in the paper illustrate this, too, for they developed only after the characteristics peculiar to the operational amplifier were recognized.

Many other examples could be quoted. The operational amplifier can become a most effective laboratory tool for students well-versed in its use. When students are encouraged to use it to simulate the systems being studied in lecture sessions, further applications often result. Those who have used the operational amplifiers in this manner fully expect the list of noncomputing applications to grow steadily.

H. M. Paynter (Massachusetts Institute of Technology, Cambridge, Mass.): This paper adds substantial lore to the growing body of practical doctrine of feedback instrumentation. Plug-in operational amplifiers are now trustworthy items of laboratory ware.

While the writer concurs enthusiastically with the examples and philosophy contained in the paper, he feels that the reader who is new to the art might appreciate additional references which extend the scope of applications and technique. Thus, he calls attention to the recent writings in the following bibliography as being of interest.

BIBLIOGRAPHY

1. LACE (THE LUTON ANALOG COMPUTING ENGINE), PART 1, R. J. Gomperts, D. W. Righton. *Electronic Engineering*, London, England, July 1957, pp. 306–12.

2. LACE (THE LUTON ANALOG COMPUTING ENGINE), PART 2, J. C. Jones, D. Readshaw. *Ibid.*, Aug. 1957, pp. 380–85.

3. A PALIMPSEST ON THE ELECTRONIC ANALOG ART (book), H. M. Paynter, editor. George A. Philbrick Researches, Inc., Boston, Mass., 1955.

4. APPLICATIONS MANUAL FOR OCTAL PLUG-IN COMPUTING AMPLIFIERS. George A. Philbrick Researches, Inc., 1956.

5. UTILITY PACKAGE AMPLIFIER. *UPA-2 Manual*, George A. Philbrick Researches, Inc., 1958.

6. ANALOG COMPUTER TECHNIQUES APPLIED TO INDUSTRIAL INSTRUMENTATION, B. Seddon. Paper presented, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., Jan. 16, 1958. Reprinted by George A. Philbrick Researches, Inc.

7. NEW INTEGRATING CIRCUIT AND ELECTRONIC ANALOG FOR TRANSIENT DIFFUSION AND FLOW, J. R. MacDonald. *The Review of Scientific Instruments*, New York, N. Y., vol. 28, no. 11, Nov. 1957, pp. 924–26.

Paul E. Pfeiffer: Each of the discussers has contributed to the aim of the paper as stated in the introduction. It is encouraging to have Mr. Cox refer to the approach as "an organized method that would lead logically into new circuits." The paper began as a survey article for a tutorial session on analog computers. In the process of organizing the material, the author discovered circuits and applications that were new to him at the time. These have contributed to several research projects, one of which is reported briefly in reference 2 of the paper and helpfully elaborated upon by Prof. Whitaker. Mr. Cox describes interesting additions to and variations of the applications reported in the paper, some of which represent his own unpublished work. Prof. Paynter provides a useful list from the rapidly expanding number of papers on various aspects of analog simulation.

Silicon-Controlled Rectifiers from Oxide-Masked Diffused Structures

R. W. ALDRICH
NONMEMBER AIEE

N. HOLONYAK, JR.
NONMEMBER AIEE

THERE is a considerable body of literature on the theoretical and experimental characteristics of 2- and 3-terminal *p-n-p-n* negative resistance switches.[1–4] This paper describes a 2-impurity simultaneous-diffusion process for producing structures in silicon which are suitable for 2- or 3-terminal signal and power *p-n-p-n* switches ("controlled rectifiers"); the properties of the devices produced by this process are discussed in the following.

In a previous paper a mechanism is described by which almost any 4-region device of *p-n-p-n*, *p-n-p-m*, or *n-p-n-m* configuration can be made to exhibit switching action.[3] It is shown that the field resulting from current flow through the interior base regions can (if thermal destruction does not occur first) increase minority carrier transport efficiency (i.e., increase alpha) across one or both base regions until the switching condition $\alpha_{NPN}+\alpha_{PNP}\geqslant 1$ is reached. Switching will not occur only if the sum of the respective emitter efficiencies $\gamma_{NPN}+\gamma_{PNP}$ is less than unity at high currents.

Device Theory and Design

A typical structure and summary of electrical characteristics is shown in Fig. 1. One exterior region, an *n*-type emitter region, is shown grounded. Adjacent to it is a thin (high-alpha) *p*-type base region to which is attached a control electrode. Next to the thin *p*-type base region is a relatively thick *n*-type base region of dimension W_N, resistivity ρ_N, minority carrier (hole) lifetime τ_p, and hole mobility μ_p. Finally there is, adjacent to the thick *n*-type base region, an exterior *p*-type region which acts as an emitter and to which is affixed one of the main current-carrying electrodes of the device. If this region were not present, the device would be a conventional *n-p-n* transistor (of relatively high alpha-α_{NPN}) connected in the grounded emitter configuration.

The expression for the current I through the collector junction between the thick *n* and thin *p* base regions is[3,4]

$$I=\frac{I_{co}'+I_b\cdot\alpha_{NPN}}{1-(\alpha_{NPN}+\alpha_{PNP})} \tag{1}$$

where I_{co} is a saturation current term with some modification due to changes in alphas and due to multiplication,[1] I_b is the control current, α_{NPN} is the electron alpha across the thin *p*-type base region, and α_{PNP} is the hole alphas across the relatively thick *n*-type base region of width W_N. It is obvious that at relatively low currents, where $\alpha_{NPN}\sim 0.7$ to 0.9 and α_{PNP} is negligible, the device behaves

---

Paper 58-1260, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 27, 1958; made available for printing August 28, 1958.

R. W. ALDRICH and N. HOLONYAK, JR., are with the General Electric Company, Syracuse, N. Y.

The authors acknowledge the assistance of C. Howard, D. Jillson, and J. Dietz in portions of this work.

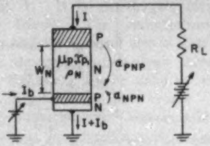


Fig. 1. Current characteristics prior to and at onset of switching of n-p-n-p controlled rectifier (switch)

$$I=\frac{I_{co}'+I_b\cdot\alpha_{NPN}}{1-(\alpha_{NP}+\alpha_{PNP}}$$

$$=\frac{I_{co}'+I_b\cdot\alpha_{NPN}}{(1-\alpha_{NPN})\left(1-\frac{\alpha_{PNP}}{1-\alpha_{NPN}}\right)}$$

$$J_B\sim\frac{W_N}{\mu_p\rho_N\tau_p\ln(1-\alpha_{NPN})^{-1}}$$

$$=\frac{W_N}{\mu_p\rho_N\tau_p\ln\beta_{NPN}'}$$

$$\text{where }\beta'_{NPN}\equiv\frac{1}{1-\alpha_{NPN}}\approx\frac{\alpha_{NPN}}{1-\alpha_{NPN}}\equiv\beta_{NPN}$$

much as a grounded emitter transistor. (The effects of multiplication are not considered; the device is generally biased well below the avalanche voltage of the collector junction. However, by properly defining the alphas to include multiplication, if any, equation 1 is still valid. In 2-terminal operation, on the other hand, avalanche or some other breakdown mechanism is required to produce the relatively high current density necessary to raise the low-voltage alpha sum to unity.) As the control current I_b is increased, both α_{NPN} and α_{PNP} increase until $\alpha_{NPN}+\alpha_{PNP}\geqslant 1$, and the device switches; i.e., the collector junction becomes forward-biased rather than reverse-biased and a total potential of the order of 1 volt exists from end to end across the device.[1-4]

α_{NPN} increases with increased current in the device because of the much-discussed trap and/or generation-recombination processes characteristic of silicon. Since the *n*-type base region is relatively thick, α_{PNP} does not increase sufficiently to produce switching by the aforementioned mechanisms. However, as I_b is increased, and a rather large electron current is caused to flow in the *n*-type base, a field is developed which causes α_{PNP} to increase to the critical value.

A theoretical expression based on a rather ideal model, which includes the effects of the field in the thick *n*-type base region, has previously been derived for the turn-on current, i.e. the load current at which the device switches to a very low forward impedance.[3] An approximate expression for the current density at which switching occurs in low lifetime material ($W_N>>L_p$, the diffusion length) is

$$J_S\sim\frac{W_N}{\mu_p\rho_N\tau_p\ln\frac{1}{1-\alpha_{NPN}}}\tag{2}$$

If α_{NPN} is relatively high, J_S is very well approximated by

$$J_S\sim\frac{W_N}{\mu_p\rho_N\tau_p\ln\beta_{NPN}}\tag{3}$$

where β_{NPN} is the grounded-emitter gain of the *n-p-n* portion of the structure. This expression contains most of the essential features of the more exact expression derived in reference 3 and is given here as useful for design purposes. Not apparent in equation 2 is the existence of a value of W_N for which $J_S=0$. Values of W_N less than this minimum result in switching without the mechanism here described. Equation 2 or 3 indicates that the turn-on current (current in the load) is directly proportional to the dimension W_N and is inversely proportional to μ_p, ρ_N, τ_p, and ln β_{NPN}. The inverse dependence upon the logarithm of β_{NPN} makes the turn-on current relatively insensitive to α_{NPN}. Fig. 2 shows experimental turn-on currents as a function of W_N in a set of devices in which area, ρ_N, τ_p, μ_p, and α_{NPN} were held constant; the turn-on currents exhibit the expected behavior as a function of W_N.

The condition that turn-on be field-dependent (and hence current-dependent) is given by

$$W_N>L_p\cosh^{-1}\frac{1}{1-\alpha_{NPN}}\tag{4}$$

Within the limits set by equation 4 considerable latitude in design is possible. The important design factor is the quantity

$$W_N/\left(\rho_N\left[\frac{q}{kT}\right]L_p^2\right)=W_N/(\rho_N\mu_p\tau_p)\tag{5}$$

W_N may be determined by practical considerations such as mechanical strength, etc. ρ_N may be determined by either the forward or reverse breakdown voltage (avalanche voltage) desired. The mobility μ_p is dependent upon the semiconductor material. Lifetime τ_p is controllable over several decades and thus is the major independent variable.

It should be mentioned that since the load current at which switching occurs is designable, this in some measure allows control over the current I_b which will produce switching. It has previously been shown[3] that the current I_b which leads to switching may be approximated by dividing the current as given by equation 2 by $\beta_{NPN}=\alpha_{NPN}\cdot(1-\alpha_{NPN})^{-1}$.

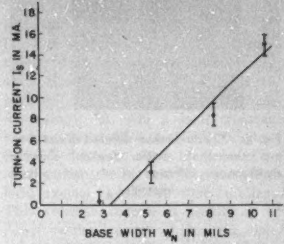


Fig. 2. Oxide-masked diffused-silicon n-p-n-p turn-on characteristics (diameter 105 mils)

If the *p*-type emitter forms a high breakdown voltage junction with the *n*-type base (W_N), it is immediately apparent that the device will exhibit a high PIV (peak inverse voltage). If an *m* (for metal) region is substituted for the *p*-type emitter, a switching structure is still possible (if $\alpha_{NPN}+\gamma_m>1$) but generally is of low PIV.

An All-Diffused P-N-P-N Structure

It is possible to produce an all-diffused structure of the type shown in Fig. 1. A process for producing high-alpha *n-p-n* structures in silicon has been previously described.[5] The diffusion is carried out in a sealed quartz tube with the concentration of two impurities (gallium and phosphorus) being controlled by the temperature and composition of an alloy source consisting of silicon, gallium, and phosphorus. This is immediately adaptable to fabrication of *n-p-n-p* switches by alloying a *p*-contact (say aluminum) to what normally would be the high-resistivity collector region. The inverse (blocking) characteristics of such a device

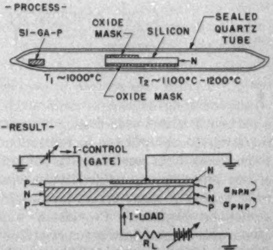


Fig. 3. Oxide-masked diffused-silicon n-p-n-p controlled rectifier

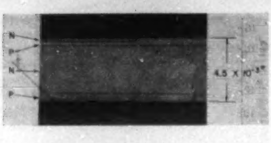


Fig. 4. Oxide-masked diffused-silicon n-p-n-p controlled rectifier (switch) showing simultaneous diffusion of n- and p-type impurities

would be determined by an alloy junction, and very good control of the alloying is required for reproducible characteristics. An alternative fabrication procedure is to oxidize one side of the silicon wafer prior to diffusion. The oxide layer effectively blocks the diffusion of phosphorus but has practically no effect on the diffusion of gallium.[6] On the oxidized side of the wafer no *n-p-n* structure is formed; rather, a *p-n* junction results. The over-all structure is thus *n-p-n-p* with the diffusion process determining the thickness of the internal *p*-region (one base) and the over-all wafer thickness determining essentially the thickness of the internal *n*-region (other base). The data in Fig. 2 were taken on structures formed in this manner on a single wafer of graduated thickness.

Contact to either of the interior regions (base regions) is required for 3-terminal operation. For the field alpha-varying mechanism to be effective, control of the current through the entire *n*-type base region is desired; this is most easily accomplished by varying the base current to the high-alpha *n-p-n* structure. Contact to the *p*-type base region can be made by alloying aluminum through the *n*-type emitter. A far superior method is to utilize again the properties of a thick oxide layer to obtain a more optimum geometry. If part of the upper surface, Fig. 3, of the silicon wafer is oxidized and subsequently the diffusion operation is performed, a surface *p*-region is produced under the oxide which is continuous with the base region of the *n-p-n* structure. A photomicrograph of a typical such structure is shown in Fig. 4. Nonpenetrating or slightly penetrating contacts may be affixed to the three separate surface regions by various well-known means. For example, gold and/or aluminum may be evaporated and alloyed to the appropriate regions, or alloys such as lead-silver may be used as desired.

Characteristics of the devices are similar to those previously described.[3] A composite illustrative characteristic is given in Fig. 5 where I_g represents control (or gate) current. In 2-terminal operation ($I_g=0$) forward breakover voltages of the order of 200–600 volts are obtained. Comparable peak inverse voltages are obtained. Turn-on currents of the order of 1 ma (milliampere) are usual for devices of 0.01-square-centimeter cross-sectional area. Control (gate) currents for turn-on are generally of the order of 0.01 to 0.1 ma for signal devices (less than 1-ampere load currents) and 0.1 to 10 ma for power devices. Forward voltage drops are very low; the voltage across W_N required to maintain α_{PNP} at the value required for switching is of the order of 0.1 volt. The total voltage drop across the device can be 0.6 volt or less in small devices. Values of 0.8 volt are observed for larger devices operating in the ampere region. Structures of 0.25-square-centimeter area have handled currents in the range of from 20 to 30 amperes.

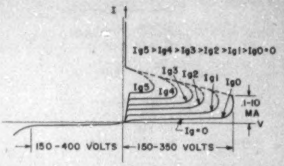


Fig. 5. Typical voltage–current and control characteristics of oxide-masked diffused-silicon n-p-n-p controlled rectifiers

References

1. P-N-P-N Transistor Switches, J. L. Moll, M. Tanenbaum, J. M. Goldey, N. Holonyak. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, Sept. 1956, pp. 1174–82.

2. P-N-P-N Switching Diodes, A. K. Jonscher. *Journal of Electronics and Control*, London, England, vol. 3, Dec. 1957, pp. 573–86.

3. Multiterminal P-N-P-N Switches, R. W. Aldrich, N. Holonyak, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 46, June 1958, pp. 1236–39.

4. The Electrical Characteristics of Silicon P-N-P-N Triodes, J. M. Mackintosh. *Ibid.*, pp. 1229–35.

5. An 85-Watt Dissipation Silicon Power Transistor, R. W. Aldrich, R. H. Lanzl, D. E. Maxwell, J. O. Percival, M. Waldner. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, vol. 5, Oct. 1958.

6. Surface Protection and Selective Masking During Diffusion in Silicon, C. J. Frosch, L. Derick. *Journal*, Electrochemical Society, Baltimore, Md., vol. 104, Sept. 1957, pp. 547–55.

Mechanization of Toll Switching at Providence, R. I.

C. W. ANDERSON
ASSOCIATE MEMBER AIEE

H. N. THORNTON
ASSOCIATE MEMBER AIEE

THIS PAPER was written to present an operating engineer's viewpoint of a particular nation-wide dialing installation from its conception to its realization. Providence, R. I., was chosen as subject of this presentation as this city appears to be typical of many others that are in service or proposed. Providence was chosen also because here is found utilization of a wide variety of the available instrumentalities and plans to incorporate probable future instrumentalities.

Previous papers have been concerned with the various requirements for nation-wide dialing as well as with major features available with various switching systems. This paper is, in a sense, a sequel to the previous papers, which were presented from a design engineer's viewpoint. This presentation is in the frame of reference of a using or application engineer.

The application engineer utilizes the selected products of the design engineer's labors to fill the peculiar combination of needs in a particular area. Forearmed with the knowledge of the requirements for nation-wide dialing, the numbering, switching, transmission, charging, signaling plans, etc., and the instrumentalities available to fulfill these requirements, he then considers the local needs.

Providence is like any other city in that its telephone needs reflect the local economy and geography and indirectly that of the neighboring area. Providence is the capitol of Rhode Island, and is located at the head of Narragansett Bay;

Paper 58-1246, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted January 9, 1958; made available for printing September 5, 1958.

C. W. Anderson and H. N. Thornton are with the New England Telephone and Telegraph Company, Providence, R. I.

see Fig. 1. It is the largest city in Rhode Island. The manufacture of jewelry and textiles and also of machine tools is important in this area. Summer tourist business is heavy along the coast line of Rhode Island and nearby Cape Cod in Massachusetts. Various military installations are significant in the area; Newport Naval Base is one and Quonset Naval Air Station another.

Telephone History of Area

The first factor to take into consideration is the telephone history of the area. The Providence toll office, to be here called Eddy Street, was installed in 1916 in leased space adjacent to a downtown local manual office. It was a no. 1 toll board and was all nondial as it has remained. This toll board was allowed to grow to meet the city's inward and through requirements and also the toll center's outward requirements. In the same leased building toll terminal facilities were installed and expanded throughout the voice-frequency era. In 1923 the conversion of local offices to dial began in the then new Washington Street building. Inward calls to this original panel office were switched under control of cordless *B* boards, assistance traffic was handled on a one-cord semimechanical board, and all outward toll calls via the Providence toll office. Nonticketed local calls were completed via panel call indicator (PCI) display at the terminating manual office.

As panel offices were added interpanel unit connections were provided by revertive pulse interoffice trunks. Handling of toll calls remained the same. Assistance calls were handled via the semimechanical board for the first two panel units and a dialing board with two cords was provided for the third unit. Two additional panel units were added in the Broad Street building under similar switching conditions except that the dial system "A" (DSA) board at that location was a type which provided key pulsing with two cords.

DSA boards had dial or key pulsing trunks to their own panel unit for local intraunit completion or through the local unit to a distant panel office with the local unit converting the call to revertive pulses. Local calls to manual office were completed also on direct trunks using PCI pulses.

At this point it is interesting to note that the two tandems now serving Providence had their beginning in about 1927. At that time Providence manual toll tandem was established as a means of providing switching access to manual toll lines from local manual and DSA boards. This manual toll tandem was a pioneer installation and was established for the purpose of completing station-to-station short-haul toll traffic and to relieve the Eddy Street outward toll board. This was the first of a series of checks on the growth of the Eddy Street toll office since outward position relief had been obtained elsewhere.

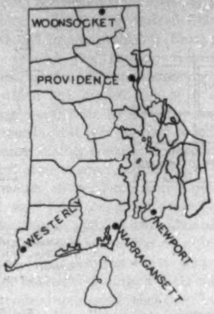


Fig. 1. Map of Rhode Island

About the same time a local manual tandem was established in the Washington Street building to serve local manual and DSA operators in reaching certain other city offices and selected nearby toll points. Inward terminal traffic from nearby toll centers was also switched through this board.

As growth and conversions continued, no. *1* crossbar was introduced into the area in Washington Street in 1940; however, as a wartime protective measure the unit was held in reserve and out of service until 1945. With the introduction of no. *1* crossbar the general method of switching remained the same except that as the unit expanded in 1948 a *3C* combined toll and DSA board was provided in Washington Street for this unit. This board handled assist traffic by d-c key pulsing, to or via its own crossbar unit to local city dial stations. Outward toll calls were switched from this board via the manual tandem toll board at the Eddy Street toll office. At that time expanded toll dialing was anticipated and the replacement of the Eddy Street toll office was probable in the not too distant future.

Shortly thereafter, no. *1* crossbar offices were provided in the Broad Street building and also in Pawtucket with similar switching arrangements and their own combined toll and DSA boards.

The five panel units were not provided with office frames. Consequently, as the interunit trunking increased either trunk expansion by means of office frames or a crossbar tandem was needed to increase the switching trunk access. Accordingly, in 1949 a local crossbar tandem was installed in the Washington Street building to meet the expanding local trunking needs and to replace the local manual tandem in that building. This crossbar tandem is essentially a 7-digit tandem capable of receiving dial-pulse, multifrequency, or revertive pulsing. Out-pulsing capabilities include revertive and dial pulsing and completion via automatic trunks and PCI. All trunks on this tandem are one-way.

Inward access to the tandem was provided from the panel and no. *1* crossbar units with revertive pulsing including office brush and group selections. The DSA operators and combined DSA and toll operators reached the tandem via multifrequency keysets (one board by dials).

A single-cord key-pulsing *B* position was provided in the Washington Street no. 3*C* switchboard lineup to provide access to the tandem from the Eddy Street toll office. Also, dial cord dialing was provided on suburban manual offices for dial-pulse access to the tandem. Completion from the tandem was via revertive pulses to the panel and no. *1* crossbar units, dial pulse to the two community dial offices (CDO). PCI pulsing was used to the metropolitan manual offices. Automatic completion was used to suburban manual offices beyond the nonticketed range.

In addition, further relief was provided to the Eddy Street toll office with the provision of a few inward multifrequency and dial-pulse terminal groups to the crossbar tandem from nearby toll centers. These nearby toll points retained ringdown groups to the Providence toll office for via and overflow terminal usage. Also provided for the use of the DSA and combined DSA and toll positions were outward dial-pulse and automatic groups to nearby toll centers for first-choice completion to that toll center central office.

Thus the development of dialing in and around Providence has been followed to the point of considerable complexity. The final step before approaching the major undertaking was the dial conversion of the suburban offices to the no. *5* crossbar. Eight suburban manual offices were so replaced between 1951 and 1956. Additional combined DSA and toll positions provided operator services with multifrequency toll switch completion

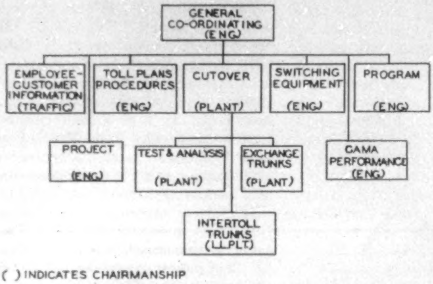

GENERAL CO-ORDINATING (ENG)
EMPLOYEE-CUSTOMER INFORMATION (TRAFFIC)
TOLL PLANS PROCEDURES (ENG)
CUTOVER (PLANT)
SWITCHING EQUIPMENT (ENG)
PROGRAM (ENG)
PROJECT (ENG)
TEST & ANALYSIS (PLANT)
EXCHANGE TRUNKS (PLANT)
CAMA PERFORMANCE (ENG)
INTERTOLL TRUNKS (LLPLT)
( ) INDICATES CHAIRMANSHIP

Fig. 2. Rhode Island tandem committee structure

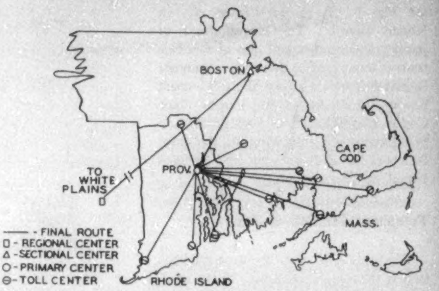


Fig. 3. Providence primary center switching plan

direct from the boards to the dial office. These offices, being suburban, relied heavily on the tandem for intracity switching.

Multifrequency pulsing was used from the no. *5* crossbar offices to the tandem and dial pulse or revertive from the tandem. Short-life investment in PCI senders was avoided either by temporary back haul via the tandem for conversion from multifrequency to PCI to the manual offices or by the interim provision of direct automatic trunks. Cordless *B* boards were avoided by the provision of the single-cord *B* position in one of the combined DSA and toll board lineups for completion of inward calls from the toll office.

In anticipation of the eventual replacement of the Eddy Street toll office, recent carrier growth had been provided in the Washington Street building and extended, voice-frequency-wise, to the toll office.

These were essentially the conditions found about 1952 when a cost study was inaugurated to investigate toll mechanization for Providence. The results of this study indicated the wisdom of replacing the toll office with a crossbar tandem equipped for centralized automatic message accounting (CAMA) operation.

Committee Organization

In order to execute such an undertaking it was felt that a formal organization of committees was needed to coordinate the various activities; see Fig. 2. In the beginning a General Coordinating Committee was organized. This committee's representation included members from all interested departments. Intercompany participation was included as most of the major cities east of the Mississippi River were involved. While this committee was large, its usefulness lay in co-ordinating the generalities of the undertaking and in following and assisting in the progress of the various working committees.

In order to pursue the switching problems, a Switching Engineering Committee was established. To co-ordinate the engineering of the toll line facilities, a Toll Plans and Procedures Committee was provided. A Programming Committee was established, and published a timetable of events for each of the various equipment orders needed by the Switching Engineering or the Toll Plans and Procedures Committees; see Fig. 2.

With this portion of the committee organization established, the project got underway. The first major consideration was whether a new tandem should be provided or the same switching accomplished by additional features and equipment on the present Providence crossbar tandem. The addition to the present tandem would cause a full tandem marker group to be reached within the decade and a new tandem would still have to be provided. Then intertandem trunking would be a problem. It was decided to provide a new toll tandem, equipped for 2-way trunking, inward and through toll service, outward operator, and CAMA switching, and also outward switching for 1- and 2-party subscribers for local automatic message accounting (LAMA) offices.

It was decided to exploit the present tandem by reducing it to an intracity tandem. It would handle base-rate coin, flat rate, and one-unit message-rate subscriber-dialed traffic, as well as intratoll center LAMA and operator-dialed traffic. The remaining capacity was used by inward terminal-grade high-usage trunks from nearby large toll centers including Boston, Mass., and New York, N. Y.

Significant decisions concerning the switching engineering are many. In the nation-wide toll switching plan, Providence had been rated as a class *III* office (primary center) homing on Boston sectional center which, in turn, homes on White Plains regional center. Homing on Providence would be all toll centers in Rhode Island, Cape Cod in Massachusetts, and toll centers in between; see Fig. 3. The noncoincident busy hours of the industrial inland areas with their day and winter busy hours offered increased efficiency in the use of toll facilities in contrast with the summer and night busy hours of the Cape, Bay, and coastal areas.

With the various offices pegged as to their position in the switching plan, the Toll Plans and Procedures Committee ascertained the intertoll trunk groups required for the undertaking.

Changes in Toll Circuit Setup

Because of the speed and efficiency of the dial machine many direct groups from the dependent toll centers to distant cities were reduced in size or eliminated. This called for a considerable change in the toll circuit setup in that facilities had to be cut off and switched at Providence.

As a portion of the toll line facilities were to be replaced at cutover and the remainder at a later data with new equipment in the Washington Street building, it was decided to extend all toll cables to the Washington Street building but to splice through only the pairs needed for cutover. Tie cables were provided for voice-frequency extension from Eddy Street carriers, etc., to the hybrids and single-frequency equipment at Washington Street. Replacement of program, teletypewriter, and the teletypewriter exchange switchboard were also included in the work.

Over-all transmission improvement was included with the introduction of via net loss (VNL) transmission treatment to the intertoll lines. Via net loss arrangements were not available for the older DSA boards. Consequently it was decided to confine their dial office assignments to dial units within the building and to operate the trunks between the DSA board and the tandem at VNL+2. For the combined DSA and toll positions, VNL connections were provided with the tandem and VNL+2 connections to their assigned offices.

Switching considerations continued. The intertoll trunks totaled 1,388. Of the 52 distant intertoll points 42 were dial, and either dial pulse or multifrequency pulsing was used in and out. For the ten manual toll centers, four were reached with automatic operation. No ringdown applique was available at first for the other six, consequently inward completion to them from the machine was planned via a Washington Street switchboard. In lieu of intertoll trunk appearance in the toll board, all dial intertoll trunks appeared on the new tandem only, which had been named the Rhode Island tandem to distinguish it from the earlier tandem called the Providence crossbar tandem.

Before the number of additional *3CL* positions could be determined, the amount of subscriber-dialed toll traffic had to be estimated. The state of Rhode Island is numbering plan area *401* and the major foreign area to these subscribers is the eastern Massachusetts area *617*. Conflicts in code assignments in these two areas had been avoided and indications were that this could be continued until after 10-digit equipment became available for panel and no. *1* crossbar offices. Plans were laid for PCI pulsing from the panel and no. *1* crossbar offices to the CAMA tandem for 7-digit calls to all available offices in these two numbering plan areas. Ten-digit dial-pulse CAMA arrangements were provided from the no. *5* crossbar offices for completion to the same offices in the *401* and *617* areas on a 7-digit basis, and on a 10-digit basis to the New York City area *212* and the Philadelphia, Pa., area *215*. Completion to nearby numbering plan areas was avoided for lack of 6-digit translation in the Rhode Island tandem as multiple entry into these areas was required. Distant numbering plan areas were also excluded for lack of adequate code treatment and 6-digit translation in the Boston sectional center which has not been converted from a no. *4* type of machine. The feeling existed that dialing should be established to those points that offered the greater volumes of dialable traffic. For the remaining points for which the availability of adequate arrangements was in question and the volumes light it was planned to establish customer dialing when arrangements permitted after cutover and in lieu of future growth in outward toll board positions. One no. *5* LAMA crossbar dependent toll center was in need of switchboard relief and it was decided to dial pulse into the Rhode Island tandem the 4- and 8-part direct distance dialing (DDD) calls and postpone switchboard relief there possibly until some local means of recording that type of traffic is available. Another dependent toll center, a step-by-step office, which was in need of switchboard relief necessitating a basement and 2-floor addition for switchboard position space, was brought into the Rhode Island tandem on a dial-pulse basis using a *112* directing code. This relief may be permanent or until a means of recording charges is provided in the step-by-step office; see Fig. 4.

Inclusion of tributaries of dependent toll centers was omitted until either charging equipment is provided at the dependent toll center or other arrangements are made. This was deemed advisable as the dialable traffic offered was not significant. One Providence remote tributary, a CDO, was omitted from the CAMA service as major toll connecting cable additions would be required and were not otherwise needed for some time. The official private branch exchange in Providence was provided CAMA service with the digit "6" as a directing code. A nonused office code "Jackson 5" plus the extension number is passed to the CAMA operator.

Completion from the tandem tributaries was provided by revertive pulsing to the panel offices and revertive or multifrequency to the no. *1* crossbar offices. Multifrequency was used to the no. *5* crossbar offices, except that provision was made for completion to certain remote offices over a group common to both tandems if required and for this dial pulse-revertive was provided. Of course dial pulse was used to the CDO's either over separate groups or groups common to both tandems. As a reserve feature two small groups of overflow trunks were provided from the Rhode Island tandem to the Providence crossbar tandem for protection in completion to the tributaries. One group equipped with 2-decibel pads provides completion through the Providence crossbar tandem to intrabuilding dial units and the other without pads for completion through the Providence crossbar tandem to offices located 2 decibels or more from the building.

With the increased efficiency offered by operator toll dialing the DSA and combined DSA and toll position could handle the needs of more stations than before. The initial DDD traffic reduced the switchboard load and still more stations could be handled by the same available positions. At the same time the 113 positions at Eddy Street were being abandoned, but with DDD and operator toll dialing they could be replaced by a lesser number of new toll positions. Combining these two conditions resulted in only 22 positions being required at Washington Street.

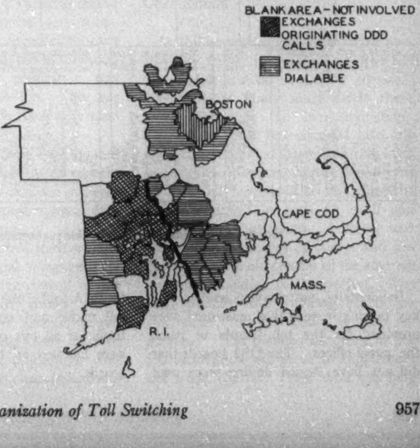


Fig. 4. Simplified DDD map

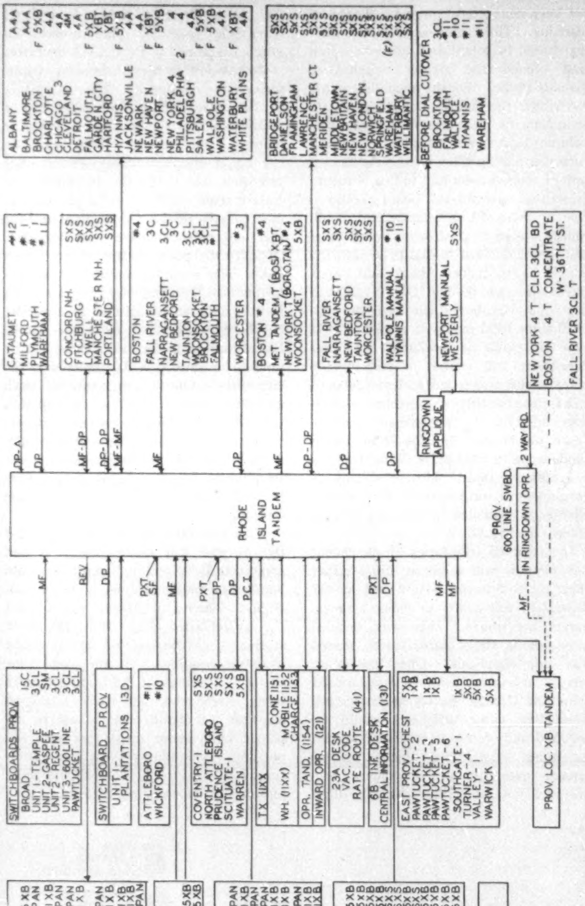

ALBANY A4A
BALTIMORE A4A
BROCKTON F 5XB
CHARLOTTE 4A
CHICAGO 3 4A
CLEVELAND 4M
DETROIT 4A
FALMOUTH F 5XB
GARDEN CITY XBT
HARTFORD XBT
HYANNIS F 5XB
JACKSONVILLE 4A
NEWARK 4A
NEW HAVEN F XBT
NEWPORT F 5XB
NEW YORK 4 4
PHILADELPHIA 2 4
PITTSBURGH 4A
SALEM F 5XB
WALPOLE F 5XB
WASHINGTON 4A
WATERBURY F XBT
WHITE PLAINS 4A
BRIDGEPORT SXS
DANIELSON SXS
FRAMINGHAM SXS
LAWRENCE SXS
MANCHESTER CT. SXS
MERIDEN SXS
MIDDLETOWN SXS
NEW BRITAIN SXS
NEW HAVEN SXS
NEW LONDON SXS
NORWICH SXS
SPRINGFIELD (F) SXS
WAREHAM SXS
WATERBURY SXS
WILLIMANTIC SXS
BEFORE DIAL CUTOVER 3CL
BROCKTON #11
FALMOUTH #10
WALPOLE #11
HYANNIS #11
WAREHAM #11
CATAUMET #12
MILFORD #1
PLYMOUTH #1
WAREHAM #11
CONCORD NH. SXS
FITCHBURG SXS
LOWELL SXS
MANCHESTER N.H. SXS
PORTLAND SXS
BOSTON #4
FALL RIVER 3C
NARRAGANSETT 3CL
NEW BEDFORD 3C
TAUNTON 3CL
WOONSOCKET 3CL
BROCKTON 3CL
FALMOUTH #11
WORCESTER #3
BOSTON #4 #4
MET TANDEM T (BOS) XBT
NEW YORK T (BORO.TAN) #4
WOONSOCKET 5XB
FALL RIVER SXS
NARRAGANSETT SXS
NEW BEDFORD SXS
TAUNTON SXS
WORCESTER SXS
WALPOLE MANUAL #10
HYANNIS MANUAL #11
NEWPORT MANUAL SXS
WESTERLY
NEW YORK 4 T CLR 3CL BD
BOSTON 4 T CONCENTRATE
W-36TH ST.
FALL RIVER 3CL T
DP-A
DP
MF-DP
DP-DP
MF-MF
MF
DP
MF
DP-DP
DP
DP
RINGDOWN APPLIQUE
2 WAY RD.
RHODE ISLAND TANDEM
PROV. 600 LINE SWBD.
IN RINGDOWN OPR.
MF
REV
DP
PXT SF
MF
PXT DP
DP
PCI
DP
PXT DP
MF
MF
MF
SWITCHBOARDS PROV.
BROAD 15C
UNIT 1-TEMPLE 3CL
UNIT 2-GASPEE SM
UNIT 2-REGENT 3CL
UNIT 3-600LINE 3CL
PAWTUCKET 3CL
SWITCHBOARD PROV.
UNIT 1-PLANTATIONS 13D
ATTLEBORO #11
WICKFORD #10
COVENTRY-1 SXS
NORTH ATTLEBORO SXS
PRUDENCE ISLAND SXS
SCITUATE-1 SXS
WARREN 5XB
TX 11XX
WH. (11XX) CONE 1151
MOBILE 1152
CHARGE 1153
OPR. TAND. (1154)
INWARD OPR. (121)
23A DESK VAC. CODE RATE ROUTE (141)
6B INF. DESK
CENTRAL INFORMATION (131)
EAST PROV-CHEST 5XB
PAWTUCKET-2 1XB
PAWTUCKET-3 1XB
PAWTUCKET-5 1XB
PAWTUCKET-6 1XB
SOUTHGATE-1 1XB
TURNER-4 5XB
VALLEY-1 5XB
WARWICK 5XB
PROV. LOC. XB TANDEM
CENTREDALE-1 5XB
ELMHURST-3 PAN
DEXTER-1 1XB
ELMHURST-1 PAN
GASPEE-1 PAN
HOPKINS-1 1XB
JACKSON-1
PLANTATIONS-1 PAN
-MANNING-1 1XB
STUART-1 1XB
TEMPLE-1 1XB
UNION-1 1XB
WILLIAMS-1 PAN
LAMA
EAST PROVIDENCE 5XB
-CHESTNUT
WOONSOCKET 5XB
CAMA
DEXTER PAN
EL-TE 1XB
GASPEE PAN
HO-WI PAN
PA 2,3 1XB
PA 5,6-SO-1 1XB
PLANTATIONS PAN
STUART 1XB
UN-JA 1XB
CAMA
CE-EL 5XB
EA PROV-CHEST 5XB
NARRAGANSETT 5XB
NEWPORT 5XB
NO.ATTLEBORO SXS
ORC. PBX 5XB
REGENT HO-3 SXS
SCITUATE SXS
VALLEY WI-4 5XB
WARREN 5XB
WOONSOCKET 5XB
PLANT TEST CODES

Fig. 5. Rhode Island tandem trunking schematic

Incidentally, switchboard senders (d-c key pulsing in revertive pulse-out) were provided for the toll boards to reach the panel offices. The DSA boards that did not have 2-cord dialing were used for DSA traffic only. The corresponding toll traffic and both the DSA and toll traffic of all out of building tributaries were handled on the no. *3C* and *3CL* boards.

With these provisions the cordless *B* boards and the *B* operation in the *3C* lineup were eliminated except for a small post cutover cleanup of miscellaneous traffic such as from the test desk to the panel offices.

The extra *3CL* positions provided for the ringdown operation mentioned previously had an interesting twist. The conversion schedule for these ringdown toll centers indicated that sufficient positions would be released by their conversion to dial to permit the replacement of the DSA semimechanical board. However, subsequent delays in the conversion program and upswings in toll traffic have made it necessary to continue the use of this switchboard for a few more months. In fact, during the engineering of the project ringdown appliques permitting direct access from the tandem to these ringdown trunks became available and were used to offset interim increases in position requirements; see Fig. 5.

In the original planning of the job the CAMA tapes were to be shipped to the existing automatic message accounting (AMA) center in Boston. However, interim increases in the load in the Boston accounting center indicated the advisability of providing an AMA center in Providence. It was decided to provide minimum equipment quantities in Providence with wiring duplicating that in Boston thereby permitting the processing of the Providence tapes in Boston should such an emergency arise.

All of this equipment required space and for this purpose 56,000 square feet of office space was provided in a new 4-story building on a company-owned lot behind the Washington Street building. This 56,000 feet provided for the needs of those displaced by the equipment which was added in the Washington Street building plus the AMA center and some personnel previously in leased space. All other equipment was added in the 8-story Washington Street building.

Additional Committee Work

With further reference to the organization setup (Fig. 2), it is seen that the two engineering committees "mold the features" of the Project and the Program Committee, setting the sequence for each segment of the work operation. Through release of complete notes on each committee's work it was possible for all to follow the project. These releases included coverage of the various engineering decisions, the list of all intertoll trunk circuits, and a project folder which outlined the schedule for each shipment order, building work, and outside cable.

During the period that the major orders were being placed at the factory, additional committees were organized to handle the cutover work and preparation for the cutover. For this purpose a Cutover Committee was organized. It is well if the chairman of this committee can devote his full time to this work and also that he attend the Program Committee meetings throughout. The Cutover Committee consists principally of the chairman of the various working committees plus others as required. As the cutover work was great in this case several cutover subcommittees were organized. The project is most sensitive to the Intertoll Trunk Subcommittee and the chairman of this committee, too, must spend most of his time on this project. From the Toll Plans and Procedures Committee the toll layout sketches provide the first look at what is ahead. The committee estimates the cutover devices that are required and confirms this estimate as the toll circuit orders are released. This Intertoll Trunk Committee must plan its various work and test operations for execution within the allotted time period. Too, this group must specify to the installer the desired sequence in which items of a staggered installation job are needed. If possible this committee may influence the arrangements of some of the detailed toll circuit layouts as they might facilitate the cutover procedures. This committee must also verify that test and manpower provisions can meet the work time objectives. This committee, like others, reported its progress up through the other committees.

The cutover phase of the Intertoll Committee's work involves preparation and co-ordination of the execution of the detailed intertoll trunk cutover. These plans, of course, must be co-ordinated through higher committees but essentially, under the direction of the Cutover Committee chairman, the Intertoll Committee must carry to service the intertoll trunk part of the job.

The Exchange Trunk Committee is much like the Intertoll Committee in function except that it is dealing with different types of plant. However, its job deals more than with mere loop facilities, for with VNL type of operation, 2-decibel pads, impedance compensators, and voice repeaters are found on many trunks.

The Project Committee, under the chairmanship of the maintenance engineer is concerned with the installation, its job progress and quality. Through its periodic reports, the weak points receive necessary attention.

The Test and Analysis Committee was used in Providence to see that training of personnel to maintain the switching machine was adequate. This committee followed the precut maintenance test to see that every possible disorder was corrected. During the cutover phase this committee was in constant contact with all of the offices involved and was always sensitive to trouble, particularly that of an interoffice nature.

Another committee which should be mentioned is the Committee for Employee and Public Information. This committee was organized while the order was in the factory and laid its plans during that period. A major newspaper announcement was made one year before the cutover and the program began. Throughout the entire program the employees were informed, before the public was, on the various phases of the job. They received their information through direct meetings, printed releases, and bulletin board announcements. The public was reached through newspapers, clubs, billboards, mail, radio, and television. Demonstration units as well as sound films were prepared for television and club use. Direct contacts were made with all heavy toll users. Measures of the effectiveness of the program were obtained through precutover customer sample survey to determine final precut soft-spot treatment. Office codes not used for precut local dialing were intercepted and instruction applied as required as a precut measure.

Extension of this program into the post cutover period is a valuable means of combatting postcutover soft spots. The teletypewriter exchange switchboard as it developed could be cut in advance, therefore, a separate committee was organized to handle the job. This was done primarily to relieve those concerned with the tandem project.

Another committee, the Performance Committee, was organized a few months before cutover. This committee gathered precutover customer performance data as a base for its operation.

For the cutover control itself a suite of offices was employed. Segments of one room were used for control of the exchange and intertoll trunks. Each had a group to follow specific phases of trunk release, test, and cutover. Elaborate paper-work procedures were established to provide clear indications of each work status and to slow down or stop any desired operation.

In one room the Performance Committee representative compiled all the data for speedy evaluation. Each operation was supervised by a person who could evaluate and analyze rather than be distracted with the actual performance operation himself.

Various members of the local and visiting staffs were posted as to the job's over-all progress in a briefing room. To this room were fed the various data for posting on bulletin boards. Other boards and booklets were available to explain the project's description.

The cutover was good. The difficulties encountered during the various phases were made easier by teamwork and co-operation, which is the way the telephone system works.

In the days and weeks following the cutover the Performance Committee met and reviewed and analyzed the data obtained from the equipment, traffic, and plant readings, operators' assistance tickets, service-observing information, and reports by the operators to the toll repair desk. From these analyses tests were indicated, troubles cleared, and customer instruction accentuated.

This task is not yet completed; improvements and additions are planned. New features, automatic number identification, 6-digit translation, 10-digit multifrequency CAMA senders are eagerly awaited. One addition is now on order and another being planned. Growth forecasts now show that in the early 1960's the marker group will be filled. Studies have begun to determine what is next.

Conclusions

In the earliest engineering phase possible the trunk list, particularly the list of intertoll trunks, should be released. The work of many must await these data. Probably the two engineering committees could be combined into one committee charged with starting the job, and the more thoroughly it performs its assigned function the smoother the road ahead will be both for itself and other groups that follow. Issuance and subsequent reissuance of charts and job description will aid greatly in the analysis and understanding of the project. The more detailed the trunking charts are the better, for this chart is "job insurance." The detailed narrative or the job description will serve to supplement the trunk charts. These data will help to avoid misunderstandings between different organizations as to equipment function, circuit type, etc. Interconnections between systems need thorough review as to functioning compatability and also to verify that both ends are provided with equipment in adequate quantity and type. Although many feel that this committee completes its work when the orders are placed on the factory, it should be pointed out that as the job progresses and changes and adjustments are made, there is need for the committee's co-ordination since changes are inevitable.

To relieve the various cutover committees, it might be advantageous to establish a cross-wiring committee to serve as timekeepers and pushers with regard to cross-wiring operations.

The characteristics of the toll facilities at Providence made the toll cutover and precutover work difficult. Through circuits at Eddy Street had to be cut off and extended to Washington Street at cutover (plus several other involved operations). Because of such possibilities, it is well that the toll layout sketches and the toll circuit orders be released well before the testboard-to-testboard test begins. This will allow for further analysis and planning of the test and cutover work. Too, this provides time to execute preliminary circuit order work which must precede this testing. If possible, equipment required to derive any new circuits should be scheduled for service well before the testboard-to-testboard period. These additional facilities will provide room for the preliminary circuit order work and also make possible the reduction of traffic releases necessary during the precutover period.

Above all, engineering personnel, and others involved, should be permitted to spend sufficient time on the project, rather than rush on to other work.

Binary Communication Feedback Systems

B. HARRIS
NONMEMBER AIEE

A. HAUPTSCHEIN
NONMEMBER AIEE

K. C. MORGAN
NONMEMBER AIEE

L. S. SCHWARTZ
MEMBER AIEE

SHANNON'S CODING theorem for a noisy channel[1] indicates that information can be transmitted as reliably as desired in the presence of noise, provided the transmitted rate does not exceed the channel capacity. In unidirectional systems the reliability may be improved by increasing the code length or the redundancy of the message.

Actually, situations seldom occur in which the communication is purely unidirectional. Instead, there is usually a return channel by which the transmitter is informed that reception is taking place, or by which the receiver replies to the transmitter. The question therefore arises whether the return channel can be employed to improve the reliability of the forward transmission with less complex equipment than that required by long codes. Feedback can improve the capacity of the forward channel only when the channel parameters are fluctuating, in which case it can, for infinite code lengths, increase the capacity by just that amount which would result from the transmitter independently having knowledge of these fluctuations. As will be seen, however, it can be quite effective in improving the reliability of the accepted message for a finite code length. Under certain conditions feedback can achieve a zero error probability with finite code lengths.

Feedback systems can be classified according to the type of communication carried by the feedback channel:

1. *Decision feedback,* in which the feedback channel is used only to report the decision of the receiver as to the acceptability of the received message.
2. *Information feedback,* in which the feedback channel is employed to report information about the received message to the transmitter, the decision to accept, or reject and correct, subsequently being made at the transmitter.

In either type of system, once a decision is made that the message cannot be interpreted, the transmitter may repeat, in whole or in part, the ambiguous message, using the same code, or by agreement the system may shift to a higher redundancy code, or both may occur. A separate classification of feedback systems can thus be made by distinguishing between fixed-code systems and variable-code systems.

The advantage of a variable-code feedback system is that the redundancy of the code can be efficiently matched to the interference. This permits the setting of an upper limit on the probability of error, whereas for a unidirectional link the redundancy must be chosen so that the error probability will not be excessive at the poorest signal-to-noise ratio which is anticipated.

The choice of the reliability parameter to be used in analyzing each of these systems should depend on the nature of the final decision process. The relative values of per-unit equivocation and error probability, for example, are ultimately determined by what might be termed the "final" decoder, i.e., the person who acts on or interprets the received information. If the act or interpretation is such that all errors in transmission are treated identically, then the error probability is the better criterion. If, however, the incorrect message is used, in conjunction with the conditional probabilities and the structure of the code, as a basis for determining the transmitted message, per-unit equivocation may be better.

Some feedback systems exhibit the property of nonconservation of message length, but the problem of determining the reliability parameter in terms of the usefulness of the system remains unchanged. The concept of per-unit equivocation can, of course, be readily extended by utilizing the conditional probabilities of message perturbation as the message length becomes indefinitely large. To extend the probability of error requires the interpretation of the error probability per word in terms of the error probability per message.

Another useful parameter in comparing the various feedback and unidirectional systems is average normalized transmission time. This is defined as the ratio of the number of digits transmitted to

---

Paper 58-1233, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted November 7, 1957; made available for printing August 25, 1958.

B. HARRIS, A. HAUPTSCHEIN, K. C. MORGAN, and L. S. SCHWARTZ are with New York University, New York, N. Y.

The research reported in this paper was performed at New York University under the sponsorship of the Electronics Research Directorate, Air Force Cambridge Research Center, Air Research and Development Command, Laurence G. Hanscom Field, Bedford, Mass.; and the Air Force Cambridge Research Center and Air Research and Development Command, Baltimore, Md., under Contract Nos. *AF 19(604)-1049* and *AF 19(604)-1964*. The authors would like to acknowledge their indebtedness to their colleague, Prof. S. S. L. Chang for his invaluable contributions to these investigations.

Vol.
Pt. 1
195

the number of digits ultimately accepted, and thus represents the price, in increased transmission time, which must be paid to realize the error-reducing capabilities of feedback.

Decision-Feedback Systems

GENERAL DECISION FEEDBACK[2]

In a decision-feedback system an attempt is made to transmit a member of the set of messages x_i by transmitting initially the corresponding member from the set of symbols x_{1i}. If the received message indicates with an acceptable error probability that a particular member of x_i originated the transmission, the received message is so interpreted and the transmitter is correspondingly informed. The transmitter then proceeds with the next message.

When, on the other hand, the received message is considered ambiguous, the transmitter, upon being so informed, replies with the corresponding symbol from the set x_{2i}. In a similar manner, if the information contained in the two received messages is sufficient, the process is terminated and the next message transmitted. If the information received is still insufficient to permit a decision, the corresponding symbol from the next set x_{3i} is sent, and the process is repeated.

In an attempt to send a single message x_i, therefore, the transmitter will transmit, as demanded by the receiver, the corresponding set of symbols (x_{1i}, x_{2i}, . . .). The process will be terminated only when the receiver indicates that total information received is sufficient to permit a decision to be made with a specified reliability. This type of system is thus seen to permit a more efficient use of redundancy than is possible with a unidirectional system, which has no knowledge of the ambiguity of the received message.

DECISION MECHANISMS

The decision to accept or reject a message is, in general, based on some scheme for distinguishing those points in sample space closely associated with specific message points from those points which are too ambiguous. Such a scheme, for example, may utilize one of the several systematic codes such as those suggested by Hamming[3] and Slepian.[4] The coding problem is then to choose for a given code length such message words, out of all possible words, that the chosen words are at least $(2k+1)$ units from each other. Received words not more than k_1 units distant from a message word are interpreted as that message word. Under these conditions, more than $(2k+1-k_1)$ errors per word must occur before one message is misinterpreted as another, and the system is capable of correcting up to k_1 errors. Although a number of systematic coding schemes have been evolved which are adaptable to this detection scheme,[5–13] the general problem remains unsolved.

An alternate decision scheme may be based on the use of null zones[14] at the detector. In its simplest form this technique establishes such a zone at the receiver that the reception of signals within the zone indicates that the a posteriori probabilities of the two transmitted states are too nearly equal to permit interpretation without excessive error. The receiver then requests repetition of this ambiguous information.

This type of reception may be extended to code words by noting that the interpretation of some of the received digits of the code words as nulls constitutes the recognition of additional points in the sample space of the received words at half-unit distances from the conventional binary code points. Such a half-unit point, if sufficiently close to a message word, may be interpreted as arising from that word. Otherwise, the received word is considered ambiguous, and the entire word is repeated. Such a system, because of the multilevel nature of its reception, is capable of being adjusted to improve the information rate of the system, even without feedback.[14]

Double-null reception, which bisects the null zone and thereby retains polarity information, permits points in the sample space of the received words between those of the single-null system. This scheme is thus capable of further improving the information rate over single-null reception. However, single-null reception has been shown, under certain conditions, to be capable of achieving about one half of the improvement in information rate theoretically attainable by increasing the number of receiver levels without limit. Double-null reception is capable of achieving only a small improvement over single-null. One advantage of the double-null scheme, however, is the comparative insensitivity to the adjustment of the null level.

THE FEEDBACK CHANNEL

Most of the analysis to date has assumed the feedback channel to be error-free. When this situation exists, the information rate and the error probability of the direct transmission channel remain unaltered if the feedback loop is opened. The important advantage of the feedback loop is that the entire message can eventually be received at the specified error probability.

The assumption that the errors in feedback can be neglected is quite reasonable under many conditions. In the first place, binary transmissions are usually composed of basic code groups, representing the symbols of some message alphabet. If only one bit of yes-no information need be transmitted over the feedback channel for each code group in the forward channel, the effective signal-to-noise ratio in the feedback channel can usually be maintained at a considerably larger value than in the forward channel. Since, for most types of noise which are of interest, error probability decreases quite rapidly with increasing signal-to-noise ratio, the error rate of the return channel can usually be kept quite small compared with that in the direct channel.

Furthermore, if the probability of acceptance is appreciably different from the probability of rejection, coding of the null locations will result in a considerable reduction of the required feedback capacity. The more precisely the probability of rejection is known, the more efficient this coding can be.

TYPES OF DECISION FEEDBACK

Many systems are designed to operate at moderate to large signal-to-noise ratios. The simplest arrangement then consists of adjusting the system for best operation at high signal-to-noise ratios, where the probability of requiring corrective information is small. The corrective information may thus be weakly encoded without excessive reduction in efficiency by employing simple repetition, and the information contained in the rejected message may correspondingly be discarded. Such an approach has been designated "discarding decision feedback."

When a large reduction in error probability is required, the number of repetitions is on the average very large, and the amount of information thrown away by a discarding system is excessive. It is better, in such circumstances, to use cumulative decision feedback. In this system the repetitions of a given message are integrated until the resultant crosses one of the null-zone boundaries, at which time the appropriate decision is registered and the transmsion of the next message requested. The analysis of this scheme, which requires techniques

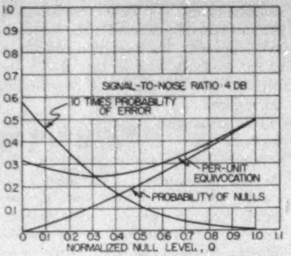


Fig. 1. Variations of parameters of binary symmetric decision-feedback (BSDF) system using single-null reception of uncoded binary pulses. Direct channel interference (DCI) is additive Gaussian noise. Feedback channel (FBC) is error-free

similar to those employed in the general random flight problem, has been worked out in some detail.[15] As expected, the method proves very powerful in reducing error probability.

When null reception is employed in a decision-feedback system, the receiver has the interesting option of terminating, or truncating, the feedback process at any time. Thus, for example, the receiver may elect, or may be forced by memory limitations, to terminate the feedback process after some specified maximum number of repetitions. In addition, the receiver also has the option of changing the null adjustment for each repetition of the same digit, either to optimize rate or to ensure a rapid convergence of the feedback process. It has been shown[16] that, for a large class of interference including additive Gaussian noise, the uniform-null nontruncated decision-feedback system is optimum, in the sense that a specified error probability can be achieved in minimum time. The effects of truncation will be discussed further in a later section.

PERFORMANCE OF NONTRUNCATED DECISION-FEEDBACK SYSTEMS

The behavior of decision feedback will be illustrated by examining the characteristics of several binary symmetric feedback systems. For simplicity the interference will be assumed to be additive Gaussian noise.

The behavior of an uncoded single-null-reception decision-feedback system is shown in Fig. 1. The parameter variations are presented as functions of the null level normalized to the noise-free received-signal amplitude. The initial improvements as the null level is increased from zero are due to the rejection of an increasing fraction of the digits which would otherwise be received in error. The important function of the feedback loop is not to increase rate, but rather to permit the entire message eventually to be received at the specified error probability. When the feedback channel is error-free, for instance, the information rate and the error probability of the direct channel remain unaltered if the feedback loop is opened. The degree of activity of the feedback channel is indicated by the probability-of-nulls relation. If the null level is raised past the optimum rate adjustment, the rate suffers as a consequence of the excessive fraction of rejections. In compensation, however, the error probability can be made arbitrarily small.

Fig. 2 shows the extent by which the rate of this system will vary when the feedback system is adjusted for operation at a given error probability. It should be noted that for a sufficiently large signal-to-noise ratio and a fixed error probability the reception reduces to that of a single-null system at the optimum rate. This optimum rate is, of course, better than that of unidirectional bilevel transmission.

For example, at a signal-to-noise ratio of 6 db (decibels) the unidirectional system has an information rate of 0.84 and an error probability of 0.06. The feedback system, adjusted for optimum rate, has an information rate which is about 4% better and an error probability which is 0.01. By introducing a sufficiently high null level the error probability can be reduced tenfold to 0.001 with a corresponding rate reduction of about 10%. To reduce the error probability a hundredfold to 0.0001 would require a rate reduction of 40%.

It can be shown that at large signal-to-noise ratios considerable improvement in the error probability can be obtained with only a small change in the null level. While the per-unit equivocation will correspondingly increase considerably, the information rate will remain approximately constant as long as the absolute value of the per-unit equivocation is low. This phenomenon is characteristic of most coding schemes, namely, that the schemes operate best when the signal-to-noise ratio is reasonably large.

The relation between error probability and transmission time for an uncoded uniform-null decision-feedback system is presented in Fig. 3. Theoretically the error probability can be made arbitrarily small at the cost of a sufficiently great increase in transmission time. Practically, of course, there are limits to the attainable signal-to-noise ratio and to the amount of time which can be allotted to transmission. These restrictions set a lower limit on the error probability.

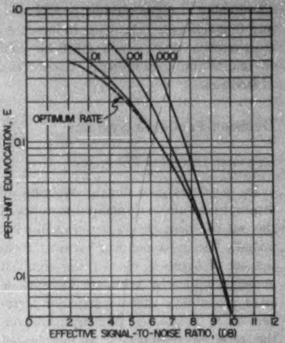


Fig. 2. Performance of uncoded BSDF system with single null adjusted for the specified error probability. DCI and FBC as in Fig. 1

A unidirectional repetition-integration system is shown in Fig. 3 for comparison, wherein each message pulse is repeated r times and integrated at the receiver before a decision is made. For convenience in analysis r is assumed to be continuous. It is seen that the feedback system is superior to the unidirectional integration system for short transmission times. For long transmission times, where the number of rejections

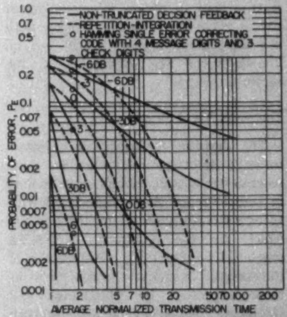


Fig. 3. Comparison of fixed-null nontruncated BSDF system with unidirectional system using Hamming error-correcting code. DCI and FBC as in Fig. 1

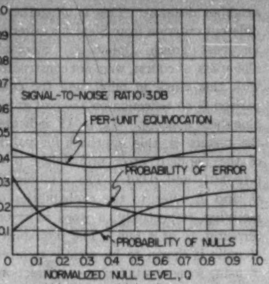


Fig. 4. Variations of parameters of BSDF system using double-null reception and 5-message-digit single-parity-check code. DCI and FBC as in Fig. 1

is large, the discarding process is very inefficient, and the integration system is better by far. If cumulative decision feedback is used, this drawback is overcome, since the feedback system then yields better performance at all values of transmission time, the asymptotic relation at large transmission times indicating a far lower probability of error than for either of the other systems.

Essentially this same comparison has been made for the case of reception in the scatter-multipath mode.[17] There the received signal is assumed to fluctuate with a Rayleigh distribution, and the null level is varied continuously in a manner which minimizes the error probability for a given transmission time. Comparison is made with a variable-integration scheme in which feedback is used to optimize the number of repetitions as a function of the received signal-to-noise ratio. Again the feedback scheme proves superior for short transmission times, and the integration scheme for longer times. Again, cumulative feedback is superior to both.

Comparison is also made in Fig. 3 with a unidirectional Hamming single-error-correcting code, operating with four message digits and three parity-check digits. The transmission time of this system is 7/4 that of the noise-free feedback system, for the same pulse rate. This system yields inferior results, as anticipated, because of its artificial constraints.

As indicated previously, binary transmissions are usually concerned, not with individual pulses, but rather with groups of pulses representing the symbols of some alphabet. The analysis must then be based on n-digit pulse groups, or words, where the per-unit equivocation and the error probabilities are calculated by noting whether the entire words, as entities, are properly interpreted and recorded.

Fig. 4 shows the behavior of a coded decision-feedback system using double-null reception and 5-message-digit single-parity-check transmission as the null level is varied. With double-null reception the interpretation of the received signal is identical to that of single-null reception except when multiple nulls are being received. Then, instead of always being rejected, the word is rejected only if parity does not check the digits indicated by the polarities of the nulls. With such a decoder the probability of error cannot be continuously reduced by increasing the null level, since very high null levels reduce the system to simple binary reception, as observed in Fig. 4. The initial effect of introducing the null level is to reduce the probability of nulls and the per-unit equivocation, and to increase the probability of error. This means that, while more words are accepted without being repeated, the probability of error of the accepted words is greater. This behavior demonstrates the previous remark that error probability and per-unit equivocation are not uniquely related to each other.

It is seen that, for a fixed signal-to-noise ratio, the ranges of variation of both the per-unit equivocation and the probability of error are limited. Also, the introduction of the null level initially increases the error probability, the value for large null levels returning asymptotically to the value for no null zone. These properties are in contrast with those of uncoded single-null reception where, by the introduction of the null, both parameters are initially improved, and where the error probability may be made arbitrarily small by increasing the null level, and hence the transmission time. A comparison may also be made with a coded single-parity-check single-null-reception system where, as the null level is increased from zero, the error probability initially increases, but ultimately becomes as small as desired, again at the cost of increased transmission time. Two obvious advantages of double-null reception over the single-null system are that the information rate for any setting of the null level is better than that of a system without nulls, and the null adjustment in the vicinity of the optimum rate setting is less critical.

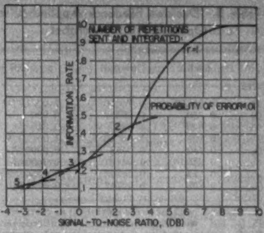


Fig. 5. Information rate versus signal-to-noise ratio for variable-code 5-message-digit single-parity-check BSDF. DCI and FBC as in Fig. 1

Fig. 5 is introduced to demonstrate the advantage of a variable code when the signal-to-noise ratio is not large. The curve is drawn for a 5-message-digit single-parity-check decision-feedback system. It is required to maintain the error probability at 0.01 or less. At low signal-to-noise ratios each digit is repeated several times by the transmitter and correspondingly integrated before decoding. For noise which is uncorrelated with the signal this results in an effective improvement in the signal-to-noise ratio, but causes a corresponding decrease in rate. The optimum arrangement, from the viewpoint

Table I. Comparison of Error Probability of Unidirectional Systems*

Information Digits	Signal-to-Noise Ratio α†	Signal-to-Noise Ratio Db	Uncoded	Hamming	Wagner	Decision Feedback
4	1.0	3.01	0.209	0.191	0.143	0.0816
	1.5	6.53	0.0349	0.0248	0.0115	0.00420
	2.0	9.03	0.00313	0.00145	0.00030	0.00008
	3.0	25.5	42×10^{-7}	6×10^{-7}	$<10^{-7}$	1.3×10^{-9}
7	1.0	3.01	0.377	0.362	0.282	0.192
	1.5	6.53	0.0789	0.0552	0.0281	0.0112
	2.0	9.03	0.00871	0.00330	0.00082	0.00022
	3.0	25.5	201×10^{-7}	12×10^{-7}	$<10^{-7}$	3.9×10^{-9}

* Employing no code, a Hamming single-error-correcting code, or a Wagner code with a decision-feedback system using a single-parity-check code and single-null reception at the optimum rate adjustment.
† This is the α of Table III in reference 18.

Table II. Decision Feedback Applied to an Optimum (10,5) Alphabet

	p=0.1	p=0.01	p=0.001
P_0	0.00063	9.7×10^{-8}	1.0×10^{-11}
Q_0	0.350	0.905	0.990
N_0	0.649	0.095	0.010
P_1	0.025	3.8×10^{-6}	4.0×10^{-8}
Q_1	0.740	0.9958	0.999955
N_1	0.235	0.0042	0.000045
P_s	0.172	0.00233	0.0000239
Q_s	0.828	0.99767	0.9999761

P_0, Q_0, N_0: with no error correction at receiver.
P_1, Q_1, N_1: with single-error correction at receiver.
P_s, Q_s: Slepian's best alphabet without feedback.

of best rate commensurate with the reliability specification, is shown in Fig. 5. For example, below a signal-to-noise ratio of about 3 db, an error probability of 0.01 can be maintained at a higher information rate by employing single repetition. Below about 0.5 db, double repetition is better, and so on.

One of the main advantages of the feedback system over a unidirectional system is that this altering of the redundancy of the code can be done as the interference level changes. Hence the higher redundancy required at low signal-to-noise ratios need not reduce the information rate at high levels.

It is of interest to compare the performance of the single-null decision-feedback system with unidirectional systems using error-correcting codes. The schemes chosen for comparison are those tabulated by Silverman and Balser.[15] These are: 1. an uncoded m-digit-word transmission, 2. a single-error-correcting code using the minimum number of parity checks, and 3. an m-digit Wagner code. The feedback code chosen for comparison is the Hamming single-parity-check code of m information digits. The null level is considered to be adjusted for optimum rate. Table I reproduces this tabulation, adding the data for the feedback system. These systems are compared on the basis of the same noise-free rates.

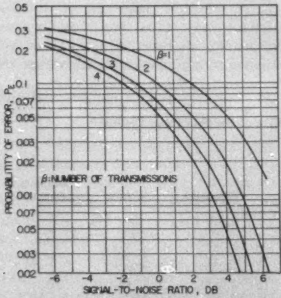


Fig. 6. Minimum possible error probability as function of maximum allowable number of transmissions per digit for uncoded BSDF system. DCI and FBC as in Fig. 1

It should be noted that, because of the rejection feature, the actual time required to transmit a message using the decision-feedback system is somewhat greater than that required by the unidirectional systems. This corresponds, over the range of m in this table, to an increase of from 23% to 40% for $\alpha=1$ and from 4% to 8% for $\alpha=1.5$. For greater signal-to-noise ratios the time devoted to repetition is negligible.

Another interesting demonstration is obtained by considering the application of Slepian's group alphabets concept to decision-feedback systems. Slepian's paper on group alphabets constitutes a significant contribution to communication theory, and his method of defining best group alphabets is elegant. One of his assumptions must be reconsidered, however, when the group alphabet concept is to be applied to feedback systems, namely, the assumption that a decision must be made on each n-group of pulses received, interpreting it as a letter of the alphabet.

In Slepian's scheme n-digit code words are sent, of which k digits carry message information. The 2^k letters of the alphabet form a group. The set of all possible received n-digit sequences are arranged in an array, with the letters of the alphabet forming the upper row. Any n-digit sequence, upon receipt, is interpreted as the letter at the top of the column of which it is a member. Each row except the top one constitutes a coset, and all elements of a coset require that the same number of digits be received in error. Slepian's best codes, essentially, are those which minimize the error probability for this type of decoding scheme.

This type of array can be utilized in a decision-feedback system. The receiver may interpret the receipt of any element of a coset as a null, or may distinguish between those cosets created by a sufficiently small number of errors and those representing several errors. In the latter instance the feedback system actually performs error correction with certain cosets, interpreting the elements of all others as nulls. It can be shown that the alphabets which minimize the error probability in this system are the so-called equal-distance codes of Hamming, i.e., those chosen to maximize the nearest-neighbor distance. This is altogether a different criterion from Slepian's criterion for best group alphabets. Slepian's alphabets, of course, were not constructed with this usage in mind, and the comparison is no reflection on the best group alphabets concept. Table II indicates the results of this approach for values of the per-digit error probability p of 0.1, 0.01, and 0.001. The symbols P, Q, and N refer to the probabilities of receiving, respectively, a word in error, correctly, or as a null. The probability of nulls is shown to be greatly reduced by using some cosets for error correction, albeit at the cost of an increase in the probability of error. The advantages of feedback are apparent in the drastic reduction in error which is achieved at the cost of a slight increase in transmission time.

PERFORMANCE OF TRUNCATED DECISION-FEEDBACK SYSTEMS

The optimum decision-feedback system employs uniform null adjustment for all repetitions, and prints a symbol only when the received signal falls outside of the null zone. With such a system there is, however, the possibility of encountering large delays before a digit is finally accepted. Such delays may exceed the memory capacity of the equipment, necessitating truncation of the sequence of repetitions. When the feedback process is truncated, it is no longer possible to obtain an arbitrarily low error probability, even at the cost of large transmission times. Fig. 6 shows the minimum attainable error probability as a function of the number

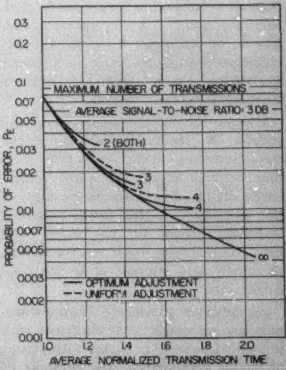


Fig. 7. Performance of uncoded truncated BSDF system. DCI and FBC as in Fig. 1

of forward transmissions β and the signal-to-noise ratio, when the interference in the direct channel is additive Gaussian.

It can be shown that in the region of large signal-to-noise ratio the introduction of truncation, with the maximum number of forward transmissions restricted to two, has the equivalent effect on the error probability of about a 2-db increase in power, as compared with the 3-db increase that results from sending each pulse twice, and integrating. The average time per message, on the other hand, is only slightly more than one half that of the integration system.

The optimally truncated system has the disadvantage that the null adjustment must be different for each repetition. Fig. 7 demonstrates that the increase in error probability which results from using a uniform null prior to truncation is small. The significant feature of truncation is the restriction on the possible reduction in error probability.

It can be demonstrated[18] that noise in the feedback channel acts substantially as truncation, i.e., it deteriorates the system and places a bound on the extent to which the transmission time can be exchanged for error probability.

Information Feedback Systems

GENERAL INFORMATION FEEDBACK[19]

The distinguishing characteristic of information-feedback systems is that all decisions as to what corrective information, if any, should be sent are made at the transmitter. The information carried by the feedback channel may be simple repetition of the received message, parity checks computed from the received message, or some more complicated code derived from the received message. By contrast, in a decision-feedback system, the feedback channel carries only yes-no information, which may be either coded or uncoded.

If the information fed back indicates that the message was probably received correctly, the transmitter proceeds with the transmission of the next part of the message. If an error is indicated, the transmitter sends a denial, or erasure, followed by corrective information. This may be done immediately, or at a later time, in coded form.

As a simple example, assume the set of message "words" is divided into several such groups that words close in message space are in different groups. The receiver indicates only the group to which the received word belongs. When the feedback process as well as the erasure information is error-free, this scheme assures that every finally accepted word is at least in the correct group. Since the coding may be so devised that the probability of receiving a wrong word from the same group is much less than that of receiving a wrong word in general, the error probability is greatly reduced.

With information feedback the transmitter has to provide additional capacity in order to keep the receiver informed, whether it is conveying new or corrective information, i.e., erasure capacity. On the other hand, the receiver gains additional information by knowing whether the original message is acceptable to the transmitter. It can be shown that, when the feedback is error-free, the gain in information at the receiver due to reduction in equivocation by the confirmation-denial or erasure, process is equal to the net information of the confirmation-denial signal.[19] Thus in these circumstances the proper use of an erasure signal does not cause a reduction in information rate.

The significant feature of information feedback, however, is not this conservation of rate, but rather the possibility of transmitting error-free messages without coding. This is possible if the feedback channel is error-free and of sufficient capacity to permit the receiver to inform the transmitter of the exact signal received via the direct channel. The transmitter can then, by means of the erasure process, correct all errors in transmission. The assumption of error-free feedback will be valid less often for information-feedback systems than for decision-feedback systems, because of the higher rates usually required of the feedback channel.

A general characteristic of feedback systems, not shared by unidirectional systems, is the possible nonconservation of message length. Whereas, for a unidirectional system the class of output messages is limited to those of the same length as the input message, for a feedback system output messages of all lengths are possible. Nonconservation of message length can occur only when the feedback channel is not error-free with regard to the distinction between the erasure word and the set of all message words.

For transmission of information which has intrinsic redundancy, such as English text, the nonconservation effect is not necessarily serious. The extraneous material and the missing material are probably detectable in the received message, if the effect is not too severe. For messages with no built-in redundancy, however, the nonconservation effect will cause dislocation of subsequent words, which may then in certain applications be considered totally in error. In such cases, the transmitted message should first be coded, to add redundancy, and information feedback used for further improvement of the reliability.

The amount of information contained in a received message is uniquely defined, irrespective of the problem of nonconservation, since there is a specific probability of the occurrence of each possible output message. However, for counting the number of errors in a received message, some artificial rule is necessary. If, for example, an error is counted each time one is made, the transmission of a single word may result in many errors. The corresponding error probability per word may therefore be greater than one. Again, in some systems where each undetected error creates a pair of message words or annihilates a pair, it may be desirable to count each such pair as one error. In some instances it may be desirable to base an error probability parameter on the exponent which describes the rate at which the probability of a message being entirely correct, as finally recorded, decreases with increasing transmitted message length.

It will become apparent that some types of information feedback have the effect of transferring part of the required system capacity from the direct channel to the feedback channel. This suggests that a compound system may be found which best utilizes the capacities available in any given communication system.

ITERATIVE DISCARDING INFORMATION FEEDBACK

An iterative discarding system is a simple information-feedback system wherein any message believed to have been received in error is corrected by erasing and then repeating. Since repetition does not make full use of the information contained in the discarded message, this simplicity of operation is obtained at the expense of rate. The feedback message need not be merely a repetition of that received from the direct channel. It may, for example, be just a parity check based on the received message.

In an iterative discarding information-feedback system, if the feedback message indicates that an erasure was incorrectly received as a message word, the transmitter sends two erasures, to erase both that message word and the preceding

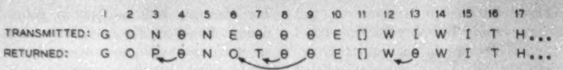

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
TRANSMITTED: G O N θ N E θ θ θ E [] W I W I T H...
RETURNED: G O P θ N O T θ θ E [] W θ W I T H...

Fig. 8. Sequence of action in transmitted message. Erasure procedure indicated by arrows

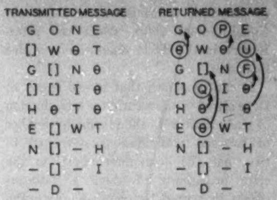


Fig. 9. Scanning of message. Circles indicate errors in reception

incorrect message. When the feedback message indicates that a message word has been incorrectly received as an erasure, the transmitter simply sends the preceding message and the intended message.

It is seen that one of the set of possible transmissions must be reserved for use as a erasure symbol. Confirmation, on the other hand, is implied merely by the transmission of subsequent information. The information that a particular message word is confirmed or denied is not entirely contained in the absence or presence of an erasure signal. To illustrate this point, an example of a discarding system with error-free and totally distinctive feedback will be given.

Consider that the message to be transmitted is GONE WITH THE WIND. Let [] denote a space. The sequence of action is shown in Fig. 8. The erasing procedure, indicated by the arrows, permits the correct message to be determined as GONE WITH. . .

From this example it may be seen that all succeeding letters serve further to confirm those already tentatively accepted. For instance, upon receiving the ninth message word, the receiver cannot be absolutely sure that an erasure was sent. If, however, some other message was wrongly received as an erasure, the transmitter would have sent an O as the tenth message instead of an E. Therefore, upon receiving the tenth message, the receiver becomes more certain that the ninth was correct. Of course, the tenth message could have been an O which was wrongly received as an E, but in that case, upon having an E fed back, the transmitter would next have sent θ instead of []. Thus, the error probability of any symbol, as evaluated at the receiver, goes to zero exponentially as the number of symbols received after the symbol in question increases. To make the error probability equal zero, one of two things is needed in addition to the error-free, totally distinctive feedback channel: an erasure word in the forward channel which cannot be transformed by errors into any other word, or an end-of-message signal that cannot be produced by errors from any other word.

One of the difficulties common to all types of feedback systems is the delay resulting from the spatial separation of the transmitter and receiver. A method of overcoming this difficulty, without stand-by operation, is to delay the correspondence between the corrective message and the original message. As an example, assume that the transmitter will receive the feedback message before transmitting the fourth succeeding letter. For convenience, arrange the message as follows:

G O N E
[ ] W I T
H [ ] T H
E [ ] W I
N D [ ]

The message will be sent by scanning each row in turn. Each column will then perform its own correcting; see Fig. 9, in which the circles indicate errors in reception. By applying the erasure rules, as indicated by the arrows, the message is recovered.

In this example the message was very short. If, for instance, the first chapter of the book *Gone With the Wind* were transmitted by this process, the number of errors in each column would be substantially equal. The time required to transmit the longest column would then be virtually equal to that required for the shortest.

In a nonfeedback system, in order to keep the coding length within reasonable limits, the rate of transmission usually must be adjusted to be appreciably less than the channel capacity. Information feedback accomplishes the same result as coding, namely, an improvement in reliability, and it also normally requires transmission at less than the channel capacity. The reasons for this loss in rate, however, are quite different from those of coding. In this case the three main causes of rate reduction are noisy feedback, rejection of the residual information contained in the erased word, and the loss of message entropy due to the existence of the erasure word.

The total signal entropy at the transmitter must equal the sum of the new message entropy and that required for the confirmation-denial process. Unfortunately, the entropy required for the confirmation-denial process is dependent on the system probabilities and hence on the signal-to-noise ratio. Thus an encoding of the system for one level of interference will result in less than optimum performance at other levels. The loss of entropy due to an unfavorable coding scheme is likely to be the most serious of the three losses when the erasure probability is high. Indeed, even with error-free feedback, this loss can result in a zero information rate even when the signal-to-noise ratio is only moderately low.

In general, however, noise in the feedback channel is the most undesirable factor in so far as rate loss is concerned. Such noise not only results in a direct loss itself but also causes the other two types of loss through an increased erasure probability. This points to the possibility of using partial information feedback to reduce the effectiveness of this noise. With less feedback information to transmit, the effective signal-to-noise ratio may be improved, with a consequent improvement in the feedback error probability.

A different situation exists when the noise in the forward channel is excessive. The direct channel transmission rate cannot be reduced without a detrimental effect on the information rate. It is possible to achieve some improvement, however, by introducing a null level and using both decision and information feedback. When in an information-feedback system the receiver is permitted the option of requesting a repeat, the feedback process then becomes compound feedback.

COMPOUND DISCARDING FEEDBACK SYSTEMS

One apparent advantage of compound feedback is that, by permitting the receiver to anticipate the transmitter, the number of erasures transmitted may be reduced. Since excessive erasure is usually the chief cause of loss in rate, compound feedback would appear to offer an improvement in this respect.

The mechanism of the compound feedback system is similar to that of an

Vol.
Pt. 1
195

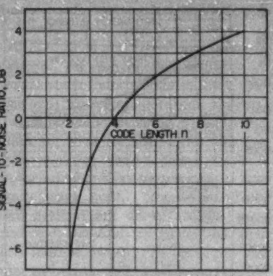


Fig. 10. Zero-information-rate signal-to-noise ratio of symmetrical-code-group repetition-feedback system. DCI and FBC as in Fig. 1

information-feedback system except that a "repeat" message is included in the set of feedback signals. The receiver sends this repeat signal back to the transmitter upon receiving an excessively ambiguous word. The establishment of a criterion of ambiguity is therefore essential to the operation of the system.

Every time a request for the repetition of a message word forestalls the use of an erasure, the transmission of an unnecessary erasure word in the direct channel is avoided. On the other hand, whenever repetition is requested of a word which would otherwise have been confirmed, an unnecessary word must be transmitted in the direct channel. An examination of the several factors which are involved indicates that, for a fairly noise-free system, the criterion of ambiguity is simply that the probability of confirmation be less than one half.

EXAMPLES OF ITERATIVE DISCARDING SYSTEMS

Simple Repetition Feedback

One of the simplest totally distinctive feedback systems is composed of symmetrical code groups, one of which is designated as the "erasure" symbol. When the feedback channel is error-free, this system permits the transfer of error-free information. The information rate is reduced over that of the corresponding unidirectional system by two effects: the use of a possible message symbol as an erasure, and the fact that no matter how an error occurs, two digits are nullified in correcting it.

At high signal-to-noise ratios the largest reduction in rate is caused by one code group being reserved as an erasure. At low signal-to-noise ratios, due to the annihilation of a pair of symbols for each error, there is a definite threshold below which the information rate becomes zero.

Fig. 10 shows the various threshold signal-to-noise ratios corresponding to a zero information rate for different code lengths and additive Gaussian noise in the direct channel. With a 5-digit code the rate becomes zero at one db. Codes having smaller lengths have lower thresholds, but then the information rate at high signal-to-noise ratios is lower because the seldom used erasure is a larger fraction of the available code groups.

One method of avoiding this threshold effect is to change the coding scheme as the signal-to-noise ratio varies. For example, a simple way of altering the transmission scheme is to employ consecutive repetition at the transmitter and integration at the receiver. The code may, in principle, be varied on a group-of-words basis or on an individual-word basis. Some direct channel capacity must be employed to keep the receiver informed about the code. The amount needed will be small if the code is not changed too often. This is the drawback which may prevent the transmitter from exercising control on a word basis.

Consider a system using 5-digit words. In the absence of repetition, the information rate is zero at a 1-db signal-to-noise ratio. By repeating each digit five times, error-free transmission becomes possible with a rate slightly less than one fifth that attainable at high signal-to-noise ratios without repetition. If the erasure procedure is not used, the information rate increases by about 5%, but the error probability becomes 0.02.

The effect of noise in the feedback channel may be partially determined by considering correct reception to occur only when the message is correctly received on a word basis. This definition neglects, however, the possibility that a sequence of events such as the

Direct channel transmission	0 Θ 0 1 1 1
Direct channel reception	0 1 1 0 Θ 1
Feedback channel transmission	
Feedback channel reception	1 Θ 0 1 1 1

may occur in an attempt to send 0111, so that even though errors are made on a word basis the final message is error-free. It is easily shown that, for small per-digit error probabilities in both channels, the error probability is proportional to the product of these per-digit error probabilities.

Table III. Complementary Feedback ($m=4$)

Direct Message				Feedback Message			
0	0	0	0	0	0	0	0
0	0	0	1	0	1	1	1
0	0	1	0	1	0	1	1
0	0	1	1	1	1	0	0
0	1	0	0	1	1	0	1
0	1	0	1	1	0	1	0
0	1	1	0	0	1	1	0
0	1	1	1	0	0	0	1
1	0	0	0	1	1	1	0
1	0	0	1	1	0	0	1
1	0	1	0	0	1	0	1
1	0	1	1	0	0	1	0
1	1	0	0	0	0	1	1
1	1	0	1	0	1	0	0
1	1	1	0	1	0	0	0
1	1	1	1	1	1	1	1

Complementary Feedback

Complementary feedback is identical to repetition feedback except that the feedback messages are not repetitions of the direct channel message. Instead, the complementary principle is employed: if two direct channel messages are close in signal space, their corresponding feedback messages are far removed, and vice versa. Specifically, if $d(x,y)$ is the distance between two direct messages x and y and $d(x',y')$ the distance between the two corresponding feedback messages x' and y', then such correspondence should be chosen that $d(x,y)$ plus $d(x',y')$ is approximately constant, or that the least sum is as great as possible.

Unfortunately, the problem of so choosing this correspondence is directly related to another unsolved problem, the problem of choosing, from 2^n symbols, such 2^m symbols that the least distance between the symbols is as great as possible. An approach to this problem may be made by examining the structure of the Hamming check-digit schemes.

One special case is that of $m=4$. In this particular case the system exhibits a neat symmetry wherein the sum of the distance between any two direct messages and the distance between the two corresponding feedback messages is always 4. In other words, for $m=4$ it is possible to pair off 2^m direct channel messages with 2^m feedback channel messages so that the sum of the distances is m, in all instances except $d(x,y)=4$, where the distance is $2m$. This arrangement is shown in Table III.

It should be noted that these feedback sequences can be indentified directly with the Hamming parity-check digits for a single-error-correcting double-error-detecting code. Such a code is, alternately, triple-error-detecting. For every m digits sent through the direct channel, c digits are fed back. Of these, (c-1)

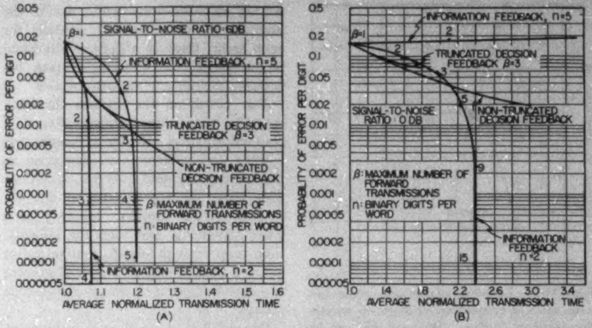


Fig. 11. Comparison of binary symmetric information-feedback and decision-feedback systems. FBC as in Fig. 1

A—When DCI is 6 db of additive Gaussian noise
B—When DCI is zero db of additive Gaussian noise

digits are used for single-error correction, the last one being a parity check over the entire set of digits in both channels.

It can be shown with complete generality that a symmetry between the direct channel and the feedback channel such as that shown here can exist only for word lengths of four or less. A difficulty arises for longer word lengths, in that the sequence of feedback digits is not always a unique indication of the received message. In fact, if there are four errors, which is the minimum distance in this code between sets of direct-channel plus feedback-channel sequences, the only cases not detected are those with the same check digits. To make identification totally distinctive, an additional identification digit must be sent over the feedback channel. Unfortunately, to make an undetected error in a word requires only four digit errors, but the number of such possibilities are reduced. Thus, although complementary feedback offers a slight improvement in some situations, it would appear that the problem of optimum coding for such an arrangement remains unsolved.

Partial Information Feedback

When the per-digit error probability in the feedback channel is not small, the probability of erasure may become excessive and greatly reduce the information rate. To minimize this effect the number of feedback digits per word may be reduced. This is then a case of partial information feedback.

There are various ways in which the feedback digits corresponding to a direct-channel word may be generated. One technique, for example, is to employ rules similar to the Hamming scheme over such feedback digits as are retained. It is difficult to state any general conclusions concerning this type of system. It accepts a reduction in the amount of information which can be fed back in exchange for an acceptable reliability of that which is fed back. As such, it offers definite improvement when the feedback channel is too noisy.

Compound Parity-Check Feedback

A partial-information system of the type just discussed may be converted into a compound feedback system by introducing a parity check into the direct channel or by employing null reception. The receiver can then, in addition to returning information concerning the received signal to the transmitter, exercise a limited power of decision regarding the acceptability of this signal. The object of this arrangement, of course, is to forestall a portion of the erasures which would otherwise have to be transmitted through the direct channel, and which have proved a major source of difficulty when the feedback is noisy. The price paid for the addition of decision feedback is that an additional "repeat" symbol must be available in the feedback channel or, alternately, one less information-feedback digit is available. Again, it is difficult to reach a firm conclusion in regard to the merits of this scheme, since no general optimizing procedure is known.

Noniterative Information Feedback

While the previous examples of iterative discarding systems can be very effective in error reduction, they suffer from one drawback, which, however, is avoidable. The requirement that every erroneous letter be erased immediately is wasteful of information rate, especially at high signal-to-noise ratios. This is due simply to the iterative mechanism, and is not inherent in information feedback, Consider, for example, an alternate system in which the transmitter sends either the entire message or some long, prearranged block of the message, meanwhile storing the information received over the return channel. At the end of this time, an erratum is transmitted which consists of coded error-position information, plus, in a coded system, corrective information. As this system minimizes the amount of information required to accomplish the erasures, the problem of a wasteful erasure mechanism is eliminated, albeit at the cost of increased storage requirements.

Truncated Information Feedback

In the nontruncated information-feedback systems previously discussed there is always the possibility of encountering large storage delays before a word is finally accepted. Since such delays may exceed the memory capacity of the equipment, there will, in a practical communication channel, be some maximum number of repetitions which cannot be exceeded. It is therefore of interest to evaluate the deterioration in system performance which results from truncation of the feedback process. The feedback channel is, for simplicity, assumed to be error-free, and the criterion of performance will be the average transmission time for a given error probability. This system is assumed to operate in the fashion of the simple, repetitive, iterative discarding system previously described.

As in noisy feedback problems, truncated information feedback results in the nonpreservation of message length. Thus the final message may contain erroneous pairs of words if an erasure is received as a word just prior to truncation, or a pair of message words may be deleted by the erroneous reception of a message word as an erasure. Because of this phenomenon it is difficult to determine word probability of error directly. It proves more convenient to calculate first the probability of receiving an entire message correctly and then to define an equivalent word probability of error as the message length is allowed to increase without limit.

In Fig. 11(A) and (B) a comparison is presented between the performance of two truncated information-feedback systems and two decision-feedback systems. The information-feedback systems are drawn for word lengths of two and five digits, and for various upper limits β on the number of transmissions of any one message word. The decision-feedback systems are drawn for the nontruncated case and for truncation after three transmissions. To facilitate the comparison, the per-word error probabilities of the information-feedback system have been recomputed in terms of equivalent per-digit error probabilities.

The information-feedback system shows a pronounced threshold effect in error probability with respect to the maximum allowable number of repetitions. The lower the signal-to-noise ratio or the larger the number of digits per word, the greater β must be before the error probability begins to decrease at a rapid rate. For values of β below this threshold the decision-feedback systems show a somewhat better performance. Above the threshold the error probability of the information-feedback system approaches zero at a rapid rate with little further increase in transmission time.

For the information-feedback systems illustrated the threshold effect occurs at about −6 db for two digits per word and at about one db for 5 digits per word. Fig. 11(B), which is drawn for zero db, demonstrates the effect of operating the 5-digits-per-word system below the threshold value.

The excellent performance of the information-feedback system above the threshold is primarily due to the control of the entire feedback process by the transmitter. It is noted that as the number of transmissions β is increased the storage requirements must be increased in order to perform the iterative discarding process. If a system having an isolated erasure digit is devised, this storage requirement can be maintained at that required for $\beta = 2$.

One method of avoiding the threshold effect, of course, is to change the coding scheme as the signal-to-noise ratio changes as, for example, by employing a suitable amount of repetition-integration with each digit.

References

1. A Mathematical Theory of Communication, C. E. Shannon. *Bell System Technical Journal*, New York, N. Y., vol. 27, 1948, pp. 379–423, 623 56.

2. Binary Symmetric Decision Feedback Systems, Bernard Harris, K. C. Morgan. *AIEE Transactions*, vol. 77, pt. I, Sept. 1958, pp. 436–43.

3. Error Detecting and Error Correcting Codes, R. W. Hamming. *Bell System Technical Journal*, vol. 29, 1950, pp. 147–60.

4. A Class of Binary Signaling Alphabets, D. Slepian. *Ibid.*, vol. 35, 1956, pp. 203–34.

5. Binary Codes with Specified Minimum Distance, M. Plotkin. *Report 15-20*, Moore Institute Research Division, University of Pennsylvania, Philadelphia, Pa., 1951.

6. A Comparison of Signalling Alphabets, E. N. Gilbert. *Bell System Technical Journal*, vol. 31, 1952, p. 504.

7. Binary Coding, M. J. E. Golay. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., vol. 4, 1954, p. 23.

8. Efficiency of Noise-Reducing Codes, A. E. Laemmel. "Communication Theory," Willis Jackson, editor. Academic Press, New York, N. Y., 1953, pp. 111–18.

9. Metric Properties of Boolean Algebra and Their Application to Switching Circuits, D. E. Muller. *Report no. 46*, Digital Computer Laboratory, University of Illinois, Urbana, Ill.

10. A Class of Multiple Error-Correcting Codes and the Decoding Scheme, I. S. Reed. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 4, 1954, p. 38.

11. Error-Free Coding, P. Elias. *Ibid.*, p. 30.

12. Coding for Noisy Channels, P. Elias. *Convention Record*, Institute of Radio Engineers, pt. 4, 1955, pp. 37–46.

13. Coding for Two Noisy Channels, P. Elias. "Information Theory. Third London Symposium," Colin Cherry, editor. Academic Press, 1955, pp. 61–76.

14. Improvement of Binary Transmission by Null-Zone Reception, F. J. Bloom, S. S. L. Chang, B. Harris, A. Hauptschein, K. C. Morgan. *Proceedings*, Institute of Radio Engineers, vol. 45, 1957, pp. 963–75.

15. Cumulative Binary Decision—Feedback Systems, S. S. L. Chang, B. Harris, K. C. Morgan. *Proceedings*, National Electronics Conference, Chicago, Ill., 1958.

16. Optimum Decision Feedback Systems, B. Harris, A. Hauptschein, L. S. Schwartz. *Convention Record*, Institute of Radio Engineers, pt. 2, 1957, pp. 3–10.

17. Binary Decision-Feedback Systems for Maintaining Reliability under Conditions of Varying Signal Strength, B. Harris, A. Hauptschein, K. C. Morgan, L. S. Schwartz. *Proceedings*, National Electronics Conference, Chicago, Ill., 1957.

18. Coding for Constant-Data-Rate Systems: Part I, R. A. Silverman, M. Balser. *Proceedings*, Institute of Radio Engineers, vol. 42, 1954, pp. 1428–35.

19. Theory of Information-Feedback Systems, S. S. L. Chang. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 2, 1956, pp. 29–40.

A 6,000-Megacycle Radio System for Toll Telephone Service

M. H. KEBBY
ASSOCIATE MEMBER AIEE

A. F. CULBERTSON
MEMBER AIEE

THE RECENTLY developed Lenkurt type *74A* microwave system is designed for a capacity of 240 telephone channels per radio circuit over short-haul routes. Economic considerations in systems planning were based upon achieving a cost per circuit-mile which would compare favorably with new toll cable installations. Fig. 1 shows two views of a basic terminal.

Equipment design includes many new techniques and components, such as miniaturization, long-life klystrons, and ferrite devices. The nominal transmitter power output is 0.5 watt and the receiver noise figure is 14 db. Operating stand-by equipment can be used to protect service against both equipment and propagation outages, and may be applied on identical or alternate frequencies to the working channels. An alternate-frequency stand-by channel may protect up to three working channels.

Power may be supplied from normal 60-cycle sources, or a transistor inverter will provide 60-cycle a-c power from telephone central office battery, where available. A highly stable transmitter automatic frequency control circuit, employing magnetic amplifiers, holds the output frequency to substantially smaller variations than are required by current FCC (Federal Communications Commission) regulations.

Applications Objectives

The type *74A* radio system was designed to be an economical transmission facility for high-density routes over relatively short distances. Basic design considerations have placed primary em-

---

Paper 58-181, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted November 8, 1957; made available for printing May 21, 1958.

M. H. Kebby and A. F. Culbertson are with the Lenkurt Electric Company, San Carlos, Calif.

Fig. 1. Two views of basic type 74A terminal

phasis on the utility of this system in telephone plant. It is also expected to find applications in private communications systems where performance standards similar to those of the common carrier industry are required.

In establishing a cost objective for this system, present-day costs per circuit mile of conventional carrier on cable facilities were taken as a basis of comparison. With the increasing use of microwave in both the Independent and Bell System plant, most transmission engineers investigate radio as a matter of course when laying out new toll routes or major expansions to existing routes. In those circumstances where freedom from right-of-way problems and freedom from closely spaced carrier repeater points are attractive, the *74A* system is expected to become a useful tool in providing a high grade of telephone service.

The 1,200-kc base band is designed to handle up to 240 channels of frequency division multiplexing such as the Lenkurt type *45BX* or similar single side-band suppressed carrier equipment. A lesser number of channels with transmitted carrier can be handled if adjustment is made for the increased carrier loading.

The system is capable of operating up to four 2-way r-f (radio-frequency) channels in parallel with transmitters and receivers on a single antenna (eight with separate transmit and receive antennas). Separation between channels is achieved by means of a ferrite circulator. A block diagram of the r-f channel multiplexing arrangements is shown in Fig. 2. Reflection phase shift across the band of signal frequencies through the two circulators has been found to be essentially negligible, even to the r-f channel traveling the longest path. The transmitter and receiver of each r-f channel are connected into a duplex arrangement by a common waveguide assembly. Separation is by means of tuned waveguide filters. A ferrite isolator between the transmitting klystron and the transmission line minimizes pulling effects. The length of waveguide which may be used between the equipment and the antenna is limited only by the loss tolerable in the transmission path.

Operating stand-by arrangements permit the use of stand-by systems on either the same or different frequency assignments. Future planning for systems with alternate-frequency stand-by contemplate up to three working channels being served by one protection channel.

Every effort has been made in the design to assure the highest order of equipment reliability. Special long-life components and tubes are used throughout. The circuit has been engineered with the revenue-producing properties of telephone traffic in mind. Wherever possible, major subassemblies are arranged on a plug-in basis. This permits ready replacement of a defective subassembly once the trouble is isolated. Maintenance test points are brought out to the front of each unit, including the i-f (intermediate-frequency) amplifier, where they are easily accessible for routine

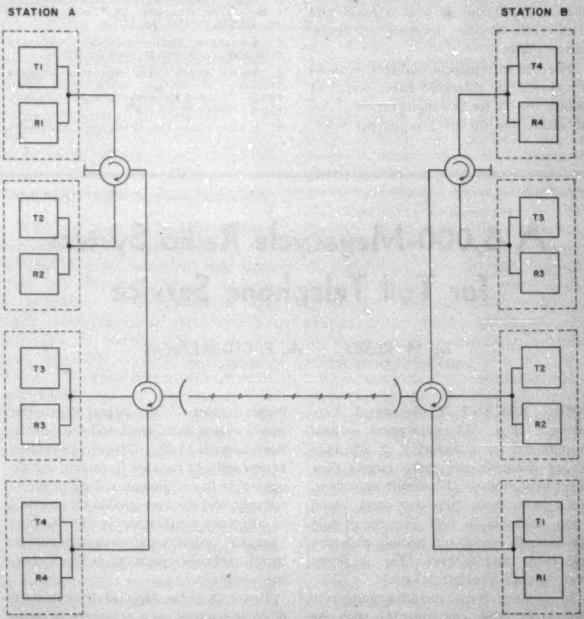


Fig. 2. Multiplexing arrangement for four transmitter-receivers on same antenna

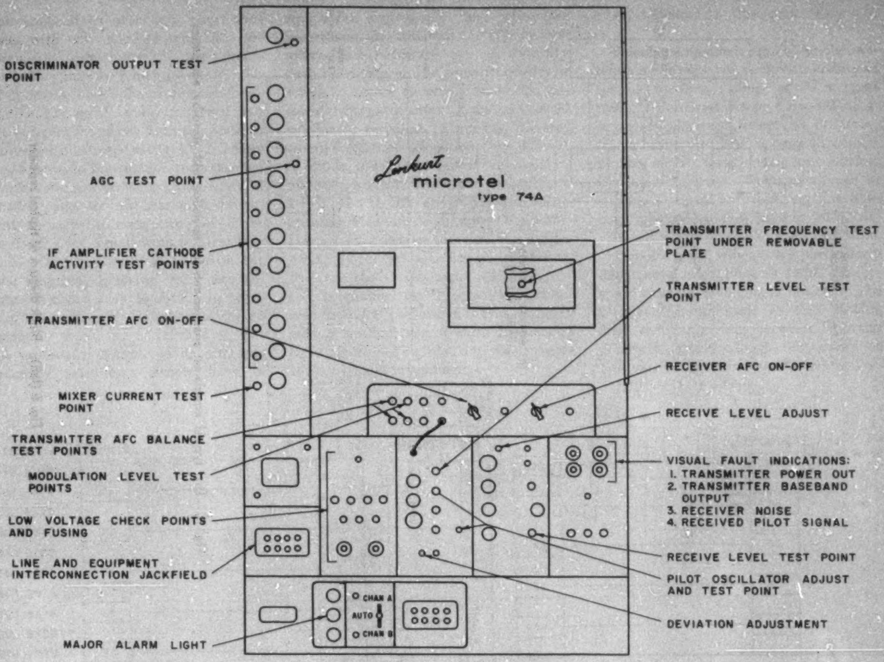

DISCRIMINATOR OUTPUT TEST POINT
AGC TEST POINT
IF AMPLIFIER CATHODE ACTIVITY TEST POINTS
TRANSMITTER AFC ON-OFF
MIXER CURRENT TEST POINT
TRANSMITTER AFC BALANCE TEST POINTS
MODULATION LEVEL TEST POINTS
LOW VOLTAGE CHECK POINTS AND FUSING
LINE AND EQUIPMENT INTERCONNECTION JACKFIELD
MAJOR ALARM LIGHT
Lenkurt microtel type 74A
TRANSMITTER FREQUENCY TEST POINT UNDER REMOVABLE PLATE
TRANSMITTER LEVEL TEST POINT
RECEIVER AFC ON-OFF
RECEIVE LEVEL ADJUST
VISUAL FAULT INDICATIONS:
1. TRANSMITTER POWER OUT
2. TRANSMITTER BASEBAND OUTPUT
3. RECEIVER NOISE
4. RECEIVED PILOT SIGNAL
RECEIVE LEVEL TEST POINT
PILOT OSCILLATOR ADJUST AND TEST POINT
DEVIATION ADJUSTMENT
CHAN A
AUTO
CHAN B

Fig. 3. Test points on basic terminal

testing. Some of the more significant test points available, in this fashion are shown in Fig. 3.

The system has been designed for maintenance with the use of conventional telephone test equipment to the fullest extent possible. Provision for operation from telephone central office battery is expected to keep outage time due to a-c power failures to a minimum.

Characteristics

Channel Noise Performance

The system is engineered to keep noise from all sources at or below a value of 31 dba (decibels adjusted) at the zero-db transmission level point under average conditions. Intermodulation noise from telephone traffic on all channels is expected to cause the total to rise above this figure for not more than 1% of the time during the busy hour. Receiver noise due to propagation conditions may be expected to increase the noise above nominal for small percentages of the time. Table I shows the amount of noise expected from various sources on a typical section of a well laid-out system.[1] The relationship between channel noise and r-f input to the receiver is shown in Fig. 4.

Operating Stand-by

For the degree of service continuity normally considered necessary in commercial toll telephone service, the use of some form of protection facility is usually good engineering practice. For the type *74A* system, this protection facility was designed to be a second operating system to safeguard against equipment failures. An important secondary benefit is the increased circuit reliability because of the second propagation path available. On

Table I. Objectives for Noise Contributions to a Typical Channel from Various Sources (Nondiversity System, Per Section)

Noise	Average Traffic and Propagation Conditions, Dba	Exceeded 1% of Busy Hour (No Fade), Dba	18-Db Fade (Condition Expected 1% of the Time), Dba	28-Db Fade (Condition Expected 0.1% of the Time), Dba
Multiplex equipment (Lenkurt type *45BX* assumed)	23	23	23	23
Modulator noise	18	18	18	18
Intermodulation noise	20	24	20	20
Receiver noise due to fades, assuming 40-db fade margin	12	12	30	40
Totals	25.8	27.2	31.3	40

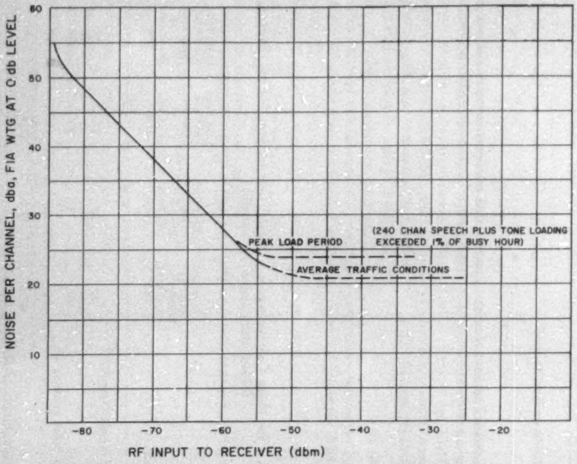


Fig. 4. Effect of radio system loading and r-f input level on channel noise for type 74A system

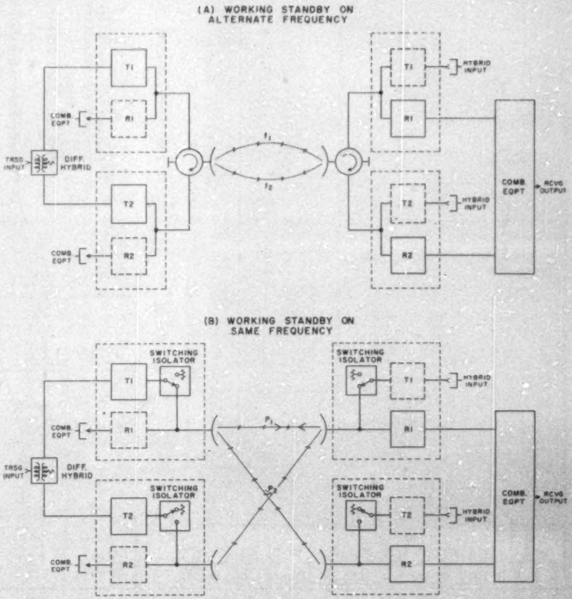

(A) WORKING STANDBY ON ALTERNATE FREQUENCY
TRSG INPUT
DIFF. HYBRID
T1
R1
T2
R2
COMB. EQPT
f_1
f_2
HYBRID INPUT
COMB. EQPT
RCVG OUTPUT
(B) WORKING STANDBY ON SAME FREQUENCY
SWITCHING ISOLATOR
P_1
P_2

Fig. 5. Two typical equipment arrangements for operation with stand-by equipment

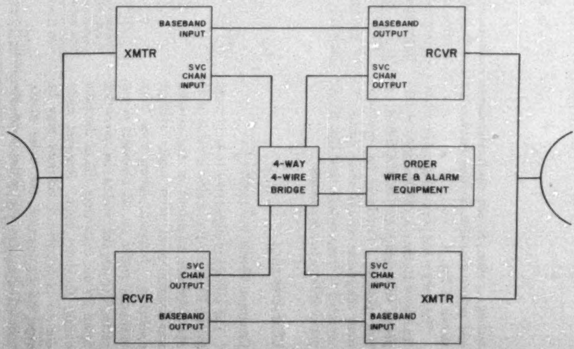

XMTR
BASEBAND INPUT
SVC CHAN INPUT
RCVR
BASEBAND OUTPUT
SVC CHAN OUTPUT
4-WAY 4-WIRE BRIDGE
ORDER WIRE & ALARM EQUIPMENT
RCVR
SVC CHAN OUTPUT
BASEBAND OUTPUT
XMTR
SVC CHAN INPUT
BASEBAND INPUT

Fig. 6 (left). Block diagram of typical repeater some applications over very short distances with only nominal reliability requirements, operation without a protection facility may be found adequate.

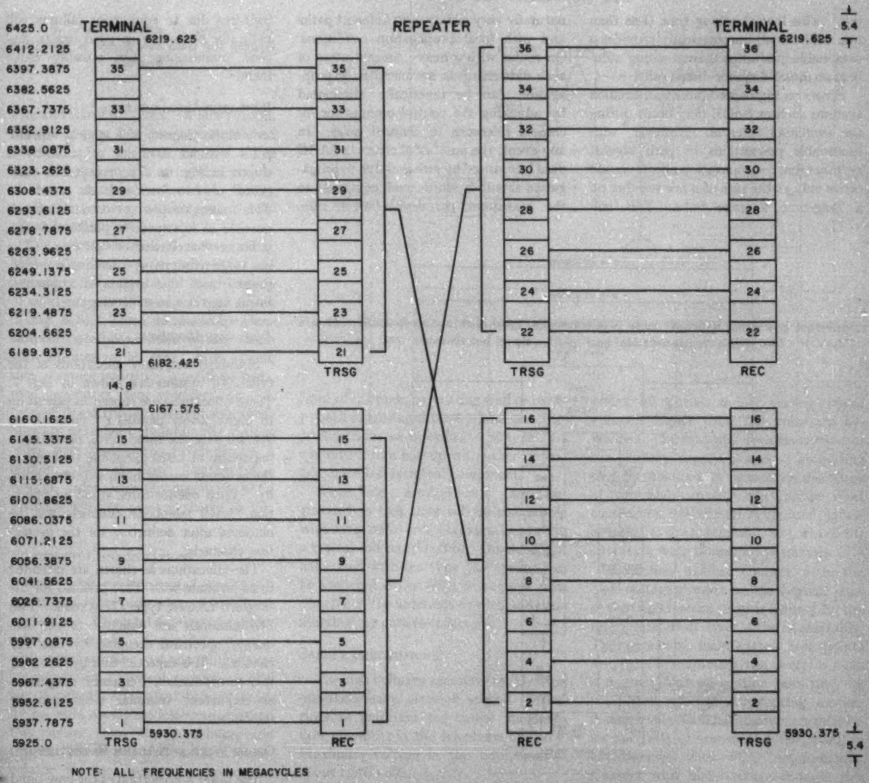
Applications with the protection channel operating on the same frequency and on a different frequency from the working channel are illustrated in Fig. 5. In either case, base-band information from the two receivers is combined in a combining and transfer panel for an average improvement in signal-to-noise ratio of 3 db (decibels) over signals from one receiver alone. Operation with stand-by on the same frequency is shown using two antennas at each station to obtain the benefit of space diversity reception of the r-f signals.

A protection channel on an alternate frequency permits a simple form of equipment arrangement and minimizes investment in antennas, towers, transmission line hardware, etc. A diversity path operating on the same frequency is somewhat more expensive to obtain because of the extra antennas and associated equipment required. However, in areas where frequency allocations are difficult to obtain, it will be preferred because of its more economical usage of the r-f spectrum. It will also be attractive on routes which grow to maximum occupancy of the available r-f channels in the common carrier band. For routes with larger circuit cross sections, the arrangement with one protection channel for three working channels will be preferable. This requires only a nominal use of extra frequencies for a protection facility, and also reduces the equipment investment.

Special Services Considerations

Telephone channels applied to the system may be expected to carry a variety of special services, along with message traffic. Foremost among these are voice-frequency telegraph, telephoto, facsimile, telemetering and SAGE (semiautomatic ground environment) data transmission. Considerable interest is being aroused in various forms of business machine data communications, and as this traffic develops it will be carried by the *74A* system.

Impulse noise, which has a particularly detrimental effect on some forms of high-speed data transmission, is expected to be relatively insignificant in this system. The switching time between diversity receivers is low enough that errors on teletype transmission should seldom oc-


TERMINAL
REPEATER
TERMINAL
6425.0
6412.2125
6397.3875
6382.5625
6367.7375
6352.9125
6338.0875
6323.2625
6308.4375
6293.6125
6278.7875
6263.9625
6249.1375
6234.3125
6219.4875
6204.6625
6189.8375
6160.1625
6145.3375
6130.5125
6115.6875
6100.8625
6086.0375
6071.2125
6056.3875
6041.5625
6026.7375
6011.9125
5997.0875
5982.2625
5967.4375
5952.6125
5937.7875
5925.0
6219.625
6182.425
14.8
6167.575
5930.375
6219.625
5930.375
5.4
5.4
35 33 31 29 27 25 23 21
15 13 11 9 7 5 3 1
36 34 32 30 28 26 24 22
16 14 12 10 8 6 4 2
TRSG
REC
TRSG
REC
REC
TRSG

NOTE: ALL FREQUENCIES IN MEGACYCLES

Fig. 7. Standard frequency assignments for type 74A radio

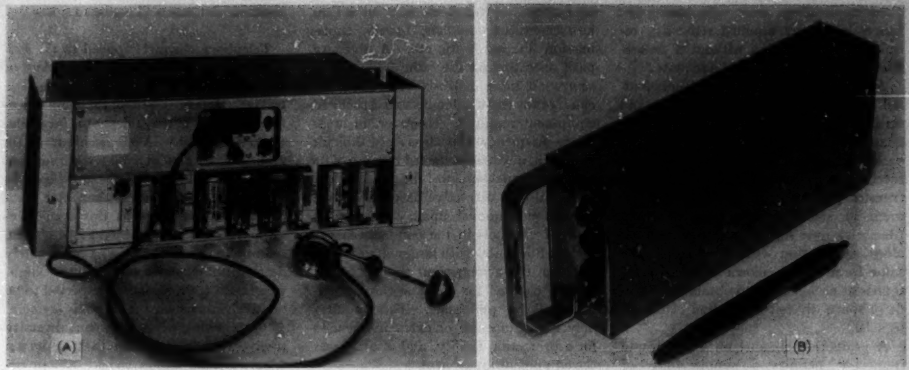


Fig. 8. A—Order wire shelf for type 74A radio. B—Typical plug-in unit for order wire shelf

cur. This low switching time (less than one millisecond) automatically provides a reasonable guarantee against wrong numbers on direct-distance-dialed calls.

Errors on high-speed data transmission systems such as SAGE may occur during the switching interval. However, with reasonable precautions in path layout, switches due to propagation effects should occur only at the rate of a few per day on a long-time average basis. This will naturally vary widely over different paths and with local propagation conditions. On routes with a heavy concentration of such data channels, switches due to propagation can be practically eliminated by adjusting the control circuits for increased tolerance to channel noise. In any event, the number of errors in SAGE data introduced by switches due to propagation should be quite small compared to the maximum permissible error rate. Switches due to equipment failures will naturally be as few as good design and good maintenance can possibly make them.

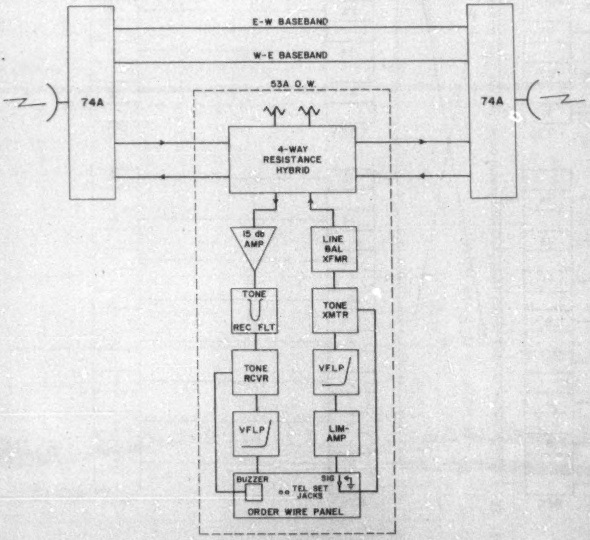


Fig. 9. Block diagram of typical order wire arrangement

REPEATER ARRANGEMENTS

A block diagram of a typical repeater point without diversity or stand-by is shown in Fig. 6. The repeater is composed of two back-to-back terminals. This makes dropout of carrier channels possible at intermediate points and simplifies service channel accessibility. The use of heterodyne or traveling-wave repeaters with this system is a possible future step in system development.

Frequency Allocations

Standard frequency allocations of the type *74A* system are shown in Fig. 7. These allocations are chosen to permit up to eight 2-way parallel r-f channels to operate over the same route, providing a maximum of 1,920 telephone channels in the common carrier band without stand-by. With one-for-three stand-by operation, 1,440 telephone channels will be obtained after deducting for the protection channels.

The allocations as chosen are intended to co-ordinate with those planned for the Western Electric type *TH* system. Two *74A* channels are assigned to the frequency spectrum occupied by one *TH* channel. It is expected that by following this co-ordination, frequency congestion at important terminal cities can be minimized.

ORDER WIRE AND ALARM FACILITIES

The lower 8-kc portion of the base band is brought to separate terminals for order

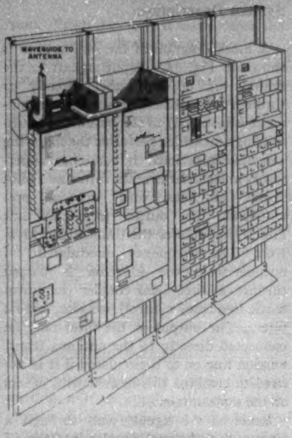


Fig. 10. Artist's sketch of typical 74A equipment in line-up with other toll terminal equipment

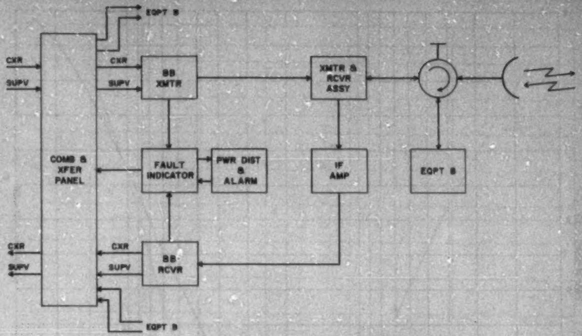


Fig. 11. Simplified block diagram of radio terminal

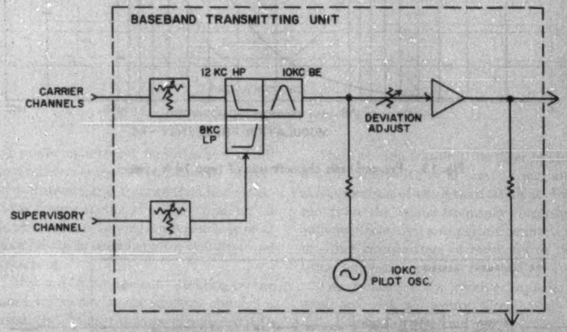


Fig. 12. Block diagram of base-band transmitting unit, which combines carrier and supervisory channels and inserts a 10-kc tone into transmission path

wire and alarm purposes. A new system, designated type *53A*, has been developed specifically for use with the *74A*. The type *53A* system is also capable of providing a universal order wire facility applicable to many other types of telephone installations.

The order wire shelf is shown in Fig. 8(A). The shelf is universally wired and is mechanically capable of mounting up to eight subassemblies of the type shown in Fig. 8(B). These mechanically identical subassemblies contain various necessary features of an order wire system such as voice-frequency amplifiers, pads, tone generators, tone receivers, etc. A total of 19 electrically different subassemblies are provided to build up the various order wire and alarm functions on a plug-in basis. A block diagram of a typical order wire arrangement at a through repeater point is shown in Fig. 9.

EQUIPMENT CONSIDERATIONS

The equipment is mounted on a standard 19-inch relay rack, with a 5-inch front projection. The total depth of the equipment is 10 inches. A complete terminal with power supplies and operating standby mounts in a vertical rack space of 9 feet, 3 inches. A terminal without standby occupies somewhat less than one half this space, since the combining and transfer panel and the waveguide circulator are not required. Access for maintenance is needed only from the front of the rack, permitting wall or back-to-back mounting of the equipment. Fig. 10 illustrates an installation of type *74A* equipment in a line-up with other typical telephone toll terminal equipment.

Waveguide arrangements are engineered so that they will be compatible with cable racks and relay rack line-ups in a typical toll central office. Outside plant associated with the type *74A* system can be arranged either with waveguide runs directly to the antennas or with antennas illuminating passive reflectors.

SAFETY PRECAUTIONS

The high voltages necessary for klystron operation have dictated special precautions to maintain the safety standards now prevalent in the telephone industry. Maximum voltage in this equipment is about 1,200 volts.

Positive-action interlocks are used under all power supply covers where voltages higher than 130 volts are developed. In addition to interruption of the primary supply, a special grounding relay is provided for immediate discharge of the high-capacity electrolytic filter condensers. Warning lights and labels are used to draw attention to parts of the circuit carrying dangerous potentials.

Protection against contact with the high voltage present on the repeller caps of the two klystron tubes is assured by the interlocked front cover of the r-f assembly. The necessary maintenance test points are accessible through the cover in such a manner that all routine tests may be conducted without interrupting service. A removable plate in the cover permits in-service tests of transmitter frequency in conformance with FCC requirements. Contact with high voltage through this opening is virtually impossible.

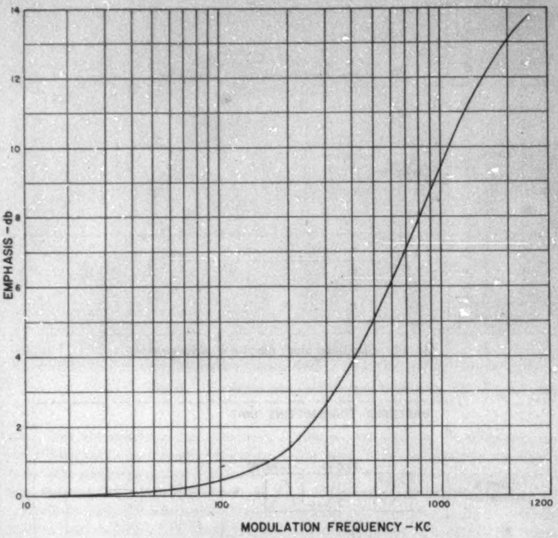


Fig. 13. Pre-emphasis characteristic of type 74A system

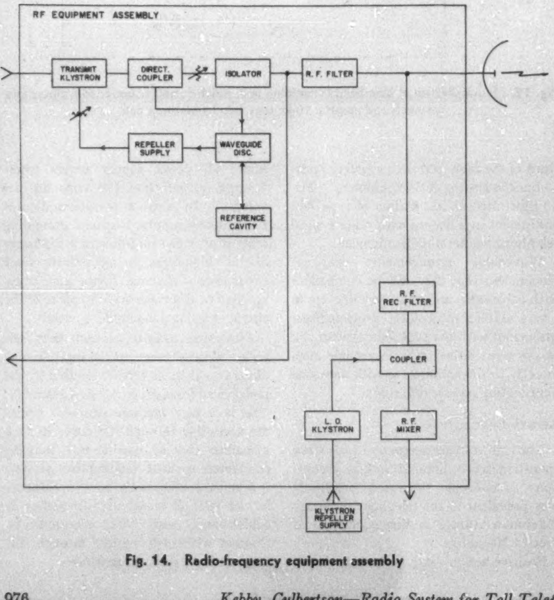


Fig. 14. Radio-frequency equipment assembly

System Description

A block diagram of the *74A* system is shown in Fig. 11. Input signals to the system are applied either directly or through a combiner (if stand by equipment is installed) to the base-band transmitting unit which performs the following principal functions: it combines supervisory and carrier circuits, furnishes a pilot signal, amplifies, and provides pre-emphasis. The output from this unit directly modulates the transmitting klystron. The frequency modulated r-f output from the klystron is passed through a directional coupler (for automatic frequency control), isolator, and r-f filter. The output of the filter may be connected directly to the antenna transmission line or to a circulator if it is desired to combine this system with others on the same antenna.

Received r-f signals pass through a filter and directional coupler into the r-f mixer. At this point the signals are mixed with a local oscillator klystron and the difference frequency of 70 mc (megacycles) applied to the i-f amplifier. An error signal from the output of the i-f discriminator is fed back to the local oscillator klystron to provide receiving automatic frequency control. The output of the i-f amplifier is amplified in the base-band receive unit which also provides de-emphasis and filtering of the carrier and supervisory circuits.

BASE-BAND TRANSMITTING UNIT

Two input connections to the base-band transmitting unit are provided: one for the telephone carrier channels and one for a supervisory channel. The carrier channel input accepts frequencies from 12 to 1,200 kc at levels from −10 to −40 db, and the supervisory channel accepts frequencies from 0.3 to 8 kc at levels from zero to −28 db. The impedance is 75 ohms for the carrier input and 600 ohms for the supervisory input. The carrier and supervisory circuits are joined through filters to provide an output base band that is continuous from 0.3 to 1,200 kc except for a 4-kc slot between 8 and 12 kc. This unoccupied portion of the base band is used to transmit a 10-kc pilot alarm tone generated in a transistorized oscillator located in this unit; see Fig. 12.

In a pure frequency-modulation system the noise in the base-band spectrum at the output of the receiver will increase at the rate of 6 db per octave unless some form of amplitude-frequency shaping (emphasis) is used. Fig. 13 shows the pre-emphasis characteristics of this system. From the curve it may be seen that the frequency deviation produced by a test tone in the top channel is about four times that of a test tone in the bottom channel. The frequency deviation is nearly constant throughout the lower portion of the base band from 12 to approximately 100 kc in order to minimize the noise generated by the klystrons. Shaping or pre-emphasis is accomplished through a resistance-capacitance network placed in the feedback path of the base-band amplifier. A bridging connection to the transmitting amplifier output is provided for sampling the 10-kc pilot to initiate alarm and transfer circuits in the event of component failure.

R-F EQUIPMENT ASSEMBLY

The output from the base-band transmitting unit is applied directly to the repeller of the transmitting reflex klystron in the r-f equipment assembly to produce a frequency-modulated output signal in the 6,000-mc band. With 240-channel busy hour loading plus signaling tones the resulting peak frequency deviation is approximately ±2 mc. The klystron has a nominal power output rating of one watt and is of the "heat sink" type; hence the need for fans and ovens is eliminated. Precise frequency control of the transmitting klystron is accomplished automatically by coupling a small portion of the output signal into a modified Pound discriminator. Comparison of the discriminator output with the reference cavity produces an error voltage which is applied to the klystron repeller to vary the frequency towards zero error. In this manner frequency stability considerably better than the present FCC requirement of 0.05% is obtained.

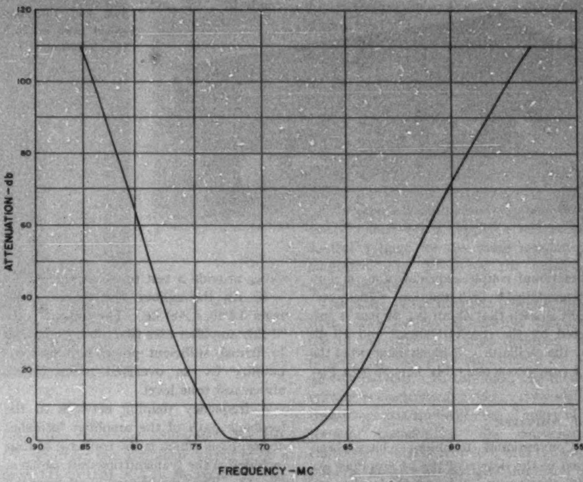


Fig. 15. Intermediate-frequency response characteristic

To provide an indication of the fade margin available in the system a variable r-f pad having a range of 30 db is provided. This variable pad may also be used to test r-f power monitoring facilities and to adjust receiver transfer equipment. To prevent antenna and transmission line reflections from causing klystron "pulling," a ferrite isolator having a reverse loss of at least 40 db is used between the filter and klystron.

If stand-by equipment operating on the same frequency as the working channel is used, the straight isolator is replaced with a switching isolator. In such cases, the stand-by transmitter is fully operative but is prevented from radiating by the 70-db reverse loss inserted by the switching isolator. This device is capable of switching in approximately one millisecond and is under the control of sensing circuits which detect loss of modulation or r-f power in the transmitter.

The 2-section transmitting filter having a 3-db bandwidth of at least 17 mc, and skirt rejection of more than 30 db at ±60 mc from the center frequency, furnishes sufficient selectivity to permit connection of other transmitters or receivers to the same antenna.

The input to the receive branch is equipped with a 4-section filter having a 3-db bandwidth of 17 mc and inserting at least 60 db loss at ±70 mc from the center frequency. This attenuation characteristic provides sufficient preselection to prevent adjacent channel interference or local oscillator radiation. In addition, it allows connection of other transmitters and receivers to a common antenna.

The local oscillator power is injected into the r-f crystal mixer through a

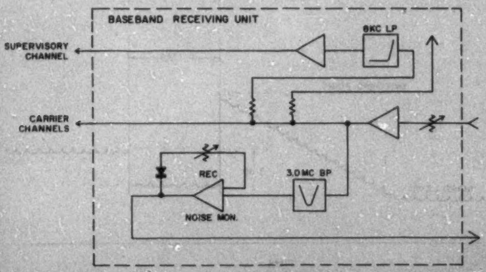


Fig. 16. Block diagram of base-band receiving unit, which separates carrier and supervisory channels and monitors out-of-band noise

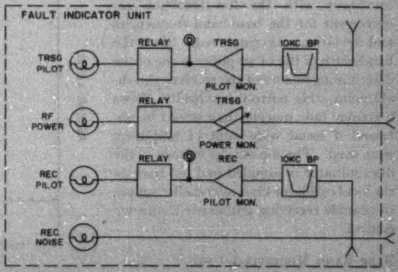


Fig. 17. Block diagram of fault indicator unit

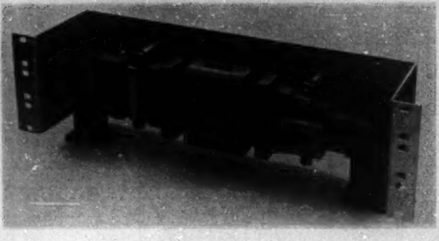


Fig. 18. Circulator panel used in 74A system

directional coupler equivalent to a nonsymmetrical hybrid junction. The power split is such that about 0.1 db loss is suffered by the received signal and 13 db by the oscillator. A block diagram of the r-f equipment assembly is shown in Fig. 14.

I-F AMPLIFER

Conventional techniques have been used in the design of the i-f amplifier except that types *408A* and *407A* tubes are used. The input stage uses the low-noise cascode design. The intermediate frequency is 70 mc and the amplifier furnishes a maximum gain of 90 db with a dynamic automatic gain control range of 60 db. This allows limiter saturation at the receiver threshold point yet insures that the amplifier will not be overloaded by a signal of up to 60 db above threshold. To insure that the gain will be maintained through tube aging, an additional 24-db gain margin is provided.

The discriminator has been designed to provide exceptional linearity over a wide band of frequencies to achieve the low degree of intermodulation necessary in a multichannel system of this type.

Typical i-f response characteristics are shown in Fig. 15. It should be noted that the bandwidth to the 3-db points is only 8 mc. This bandwidth has been selected to provide optimum performance consistent with the low-distortion requirement for the base-band frequencies and deviation ratio employed. Since the bandwidth of the i-f amplifier is one of the determining factors of the receiver threshold point, this narrow bandwidth allows a greater fade margin than would be obtained if usual wide-band i-f amplifiers were used. The d-c output voltage of the discriminator is sampled and fed back to the local oscillator klystron repeller supply to provide receiving automatic frequency control.

BASE-BAND RECEIVING UNIT

In the base-band receiving unit, three stages of amplification using type *408A* tubes provide a test tone output level of −15 db throughout the carrier band from 12 to 1,200 kc. The output stage of this amplifier uses two tubes in parallel to furnish sufficient power handling capability for an overload point 27 db above test tone level.

A frequency shaping network in the feedback path of the amplifier furnishes de-emphasis which is the converse of that included in the transmitting unit shown in Fig. 13. The supervisory band from 0.3 to 8 kc is filtered and amplified in a single transistorized stage to an output level of −4 db.

The 10-kc pilot generated in the transmitting unit is bridged off and fed to the fault indicator unit. In addition, noise at about 3 mc is filtered, rectified, and fed to the transfer and fault indicator units. Thus, both receiver noise and base-band pilot are monitored to furnish a complete indication of circuit condition. The base-band receiving unit is shown in Fig. 16.

FAULT INDICATOR UNIT

As shown in Fig. 17, the fault indicator unit contains alarm circuits to indicate failure of any active circuit in the system. The following points are monitored: the 10-kc pilot at the output of the transmitting and receiving base-band amplifiers, the transmitting klystron power output, and the receiver noise. An alarm condition lights a lamp on the face of the unit, actuates remote alarms if desired and initiates transfer to a stand-by system if equipped.

CIRCULATOR PANEL

The waveguide ferrite circulator panel has been designed to facilitate connection of more than one transmitter-receiver unit to a common antenna. It has an insertion loss of less than 0.5 db between ports and a reverse loss of more than 20 db. A single panel provides for the connection of up to three transmitter-receiver systems on one antenna; if desired, circulators may be cascaded to connect even more units to the same antenna provided the additional insertion loss is acceptable. The circulator panel is shown in Fig. 18.

COMBINING AND TRANSFER PANEL

When the system design provides for the use of stand-by equipment, both transmitters are modulated simultaneously with the carrier and supervisory information. Hybrid transformers located in the combining and transfer panel are furnished to provide this function. For this method of operation the receiver outputs are combined in a manner to allow the information content of the two receivers to add in phase or on a voltage basis; however, the noise adds on a random or power basis. Thus the signal-to-noise ratio through the combiner is

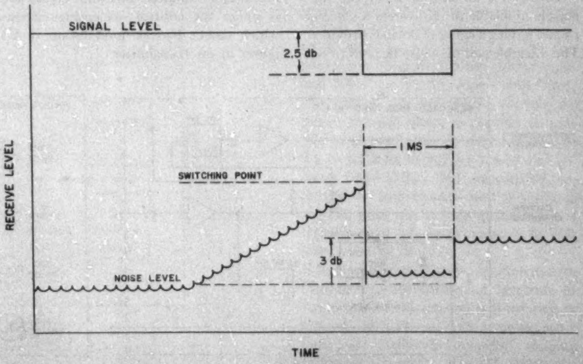


Fig. 19. Noise levels before, during, and after switching from one transmitter-receiver to another

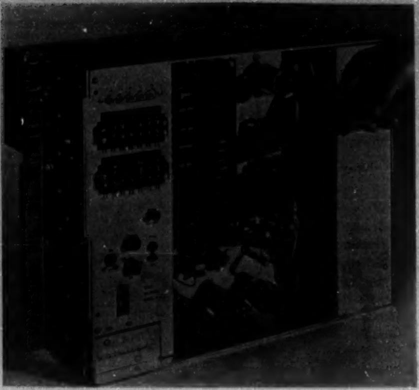


Fig. 20. Klystron power supply

improved up to 3 db over that of a single receiver when the signal strength to the two receivers is the same. Further, a decrease in intermodulation distortion products by an average of 3 db also results, since the distortion created in the two different systems is generally random in nature and not in phase.

When one of the receivers fades or fails, such as to allow the circuit noise to rise above some predetermined value, relays are actuated to disconnect and mute the noisy receiver. A control on the front of the combining and transfer panel permits adjustment of the switching point from a channel noise of 37 dba upwards. Thus, if high-speed data are transmitted over the system, it may be desirable to adjust the transfer point at 45 to 50 dba to minimize the number of switches that might occur during abnormal propagation conditions.

In the case of noise due to fading or upon forced maintenance switching, it takes less than one millisecond for the transfer from the combined output of two receivers to that of a single receiver. In the event of complete failure of one of the systems, approximately 35 milliseconds is required to silence any noise which may be present in the faulty equipment and to restore the signal level to normal.

During the switching interval, the signal level is reduced 2.5 db. After switching, the level is restored. However the signal-to-noise ratio is reduced by 3 db. This action is illustrated in Fig. 19.

POWER SUPPLIES

The system has been designed to take full advantage of the reliable source of power generally available in a central office. All the circuits using active elements, except the klystrons, have been designed for operation from 24 or 48 and 130 volts d-c.

Power for the klystrons is furnished from a supply which provides 6.3 volts d-c for the heaters and up to −1,200 volts d-c magnetically regulated for the various repeller and beam elements. In addition, magnetically controlled circuits furnish automatic frequency control for the transmitting klystron. The unit is designed for operation from 115 volts a-c, 50 to 60 cycles per second, either sinusoidal or square wave, with special care taken to minimize the effect of sharp transient variations in the supply main. All rectifiers are silicon diodes; three type *OD3* vacuum tubes supply a referenced control voltage for the main regulating amplifier. Fig. 20 shows the klystron power supply.

Operation of the klystron power supply from office battery is possible through use of an inverter panel designed to furnish 115 volts a-c from either a 24- or 48-volt battery. In this unit, the direct current is electronically chopped by eight transistors and applied to the primary of a toroidal transformer. The conversion efficiency is in excess of 90% and the output is a square wave with an amplitude of approximately 115 volts at a load of up to 200 volt-amperes.

At locations where 130 volts d-c is not available, a supply has been designed to operate from 115 volts a-c and furnish 130 volts d-c at up to 200 milliamperes with less than 10 parts per million of 120-cycle ripple. If only 24- or 48-volt d-c batteries are installed a second inverter of the type described in the foregoing may be used to supply the 130 volts d-c unit; thus, operation of the entire system is possible from only 24 or 48 volts d-c.

References

1. **NOISE CONSIDERATIONS ON TOLL TELEPHONE MICROWAVE RADIO SYSTEMS, T. A. Combellick, M. E. Ferguson. *AIEE Transactions*, vol. 74, pt. I, Mar. 1957, pp. 67–70.**

2. **TELEPHONE CARRIER FREQUENCY NETWORKS, G. D. Wallenstein. *AIEE Transactions*, vol. 73, pt. I, 1956(Feb. 1957 section), pp. 691–67.**

The Diode Reactance Modulator

G. FRANKLIN MONTGOMERY
NONMEMBER AIEE

Synopsis: The reactance modulator that uses one or two diodes to control the current in a capacitor is analyzed with the assumption of lossless circuit elements. Universal curves of susceptance, conductance, and control current, as functions of control voltage, are calculated for three basic circuits.

THE COMBINATION of two diodes and a capacitor to form a variable reactance was described as early as 1940;[1–3] more recently a single-diode circuit has been used commercially and by amateurs.[4] Such circuits are often used as frequency modulators, phase modulators, and automatic frequency controls. They are simple and convenient to apply, especially with semiconductor diodes, and apparently no detailed analysis of their operation has been published. It is the purpose of this paper to supply one.

Single-Diode Modulator

Consider the circuit of Fig. 1, where e is a sinusoidal voltage across which the capacitor C and diode D are connected in series, and E_B is a voltage that controls a direct current I. In practice, the source e is usually either the resonant circuit of an oscillator whose frequency is to be modulated, or the tuned circuit of an amplifier whose output is to be modulated in phase. E_B represents the instantaneous voltage of the modulating function; it is shown here as a battery without loss of generality, because the modulating function in any practical case varies slowly compared with the sinusoidal variation of e.

One is tempted at first to regard the diode simply as a resistor that is variable with I(or with E_B) and to proceed by referring to the known results of resistance tuning. Except for the trivial case where e is so small that the diode dynamic resistance can be considered constant throughout one cycle of e, it is obvious that the diode current is generally nonsinusoidal. The resistive assumption is therefore invalid.

It is more useful to regard the diode as a switch that is closed during some fraction of the input cycle and open during the remainder of the cycle, the period of conduction depending upon the current I and the other parameters. It may be expected that when the diode conducts throughout the cycle, the source is shunted by the capacitor C. When the diode remains nonconducting, the source is open-circuited, provided L is sufficiently large. For conduction during a fraction of the cycle, the impedance in shunt with the source should lie between these extremes.

Assume that the diode is ideal in the sense that it has zero forward resistance and zero reverse conductance. The circuit input voltage may then be defined as

$$e = E \sin \varphi = E \sin \omega t \qquad (1)$$

and an angle of diode conduction

$$\theta = \varphi_1 - \varphi_0 \geqslant 0 \qquad (2)$$

where φ_0 is the phase at which conduction begins, and φ_1 is the phase at which conduction ends. The inductor L is defined as one having infinite reactance at all angular frequencies equal to or greater than ω and zero resistance to direct current. With these assumptions, two statements may be made concerning Fig. 1:

A. During one period $2\pi/\omega$, no charge is accumulated by capacitor C.

B. During one period $2\pi/\omega$, the energy consumed by the load E_B equals the energy delivered by the generator e.

The equations equivalent to these statements are given in the Appendix. The solutions yield the ratios E_B/E and $I/\omega CE$ as functions of the angle of conduction θ, and plots of these functions are shown in Fig. 2. Note that as E_B/E decreases from unity to zero, θ increases from zero to 2π, and $I/\omega CE$ increases from zero to unity.

Of primary interest, however, is the fundamental component of current delivered by the generator. The total generator current i is sketched in Fig. 3, for $E_B/E = 0.171$, $\theta = \pi$. The fundamental component is, in general,

$$i_\omega = I_Q \cos \varphi + I_R \sin \varphi \qquad (10)$$

where I_Q and I_R are the amplitudes of the reactive and resistive components respectively. These are solved for in the Appendix. The fractional susceptance may then be defined as

$$B/\omega C = I_Q/\omega CE$$

and the fractional conductance

$$G/\omega C = I_R/\omega CE$$

These are plotted as functions of E_B/E in Fig. 4. The d-c term $I/\omega CE$ is plotted for convenience on the same graph.

The susceptance is evidently not a linear function of the control voltage E_B but has an approximately linear range for $0.15 < E_B/E < 0.25$. Determination of the exact point where $d^2B/dE_B^2 = 0$ is a problem that has not been solved, but it is convenient that this point occurs within a range where the Q of the admittance is reasonably large (about 3.7).

The circuit of Fig. 1 is directly useful in practice, but perhaps more general use has been made of the circuit of Fig. 5, in which the inductor of Fig. 1 is replaced by a resistor R large enough that its a-c conductance can be neglected. Note that E_C is shown with polarity opposite to that of E_B. The variation of susceptance with control voltage E_C is best solved for graphically, as illustrated in Fig. 6. A load line is drawn with slope $1/\omega CR$. The intersection of this line with $I/\omega CE$ represents the d-c operating

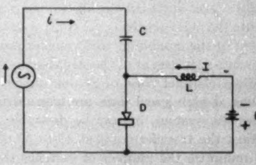


Fig. 1. Basic single-diode reactance modulator

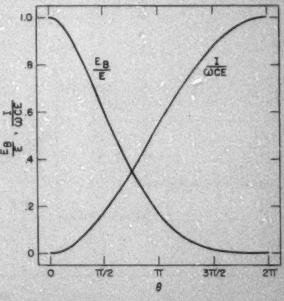


Fig. 2. Variation of control voltage and current with angle of conduction, single diode

---

Paper 58-1326, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted December 18, 1956; made available for printing September 4, 1958.

G. Franklin Montgomery is with the National Bureau of Standards, Washington, D. C.

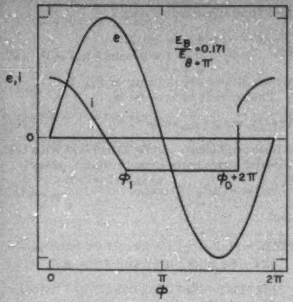


Fig. 3. Typical generator voltage and current, single diode

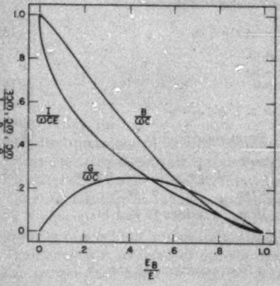


Fig. 4. Variation of susceptance, conductance, and control current, with control voltage, single diode

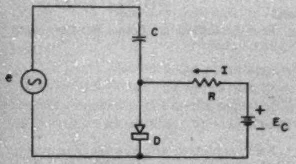


Fig. 5. Single-diode modulator with resistive control source

point for $E_C/E=0$, and the fractional susceptance is found by projecting this point vertically onto the curve for $B/\omega C$, at p_0'. For other currents, $E_C/E=q-p$, and the fractional susceptance is the vertical projection of p on $B/\omega C$, at p'. Note that the linear range of susceptance variation with E_C is now centered at $E_B/E\approx 0.25$.

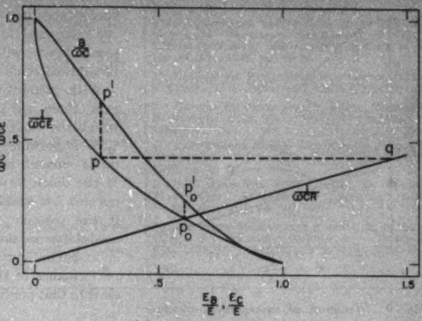


Fig. 6 (right). Graphic analysis of single-diode modulator with resistive control source

Dual-Diode Modulators

For the parallel-diode circuit of Fig. 7, definitions 1 and 2 again apply, with the qualifications that $\theta \leqslant \pi$ and that φ_0 and φ_1 are the limits of conduction for diode $D1$. Because of circuit symmetry, only the half-period $\varphi_0<\varphi<\varphi_0+\pi$ need be considered; during the remainder of the period, current i is a repetition of the first half-period current but opposite in sign. I_{DC} is defined as the direct component of current in the load E_B. In the Appendix, it is shown that $\varphi_1=\pi/2$, and with φ_1 determined, the only additional statement necessary for the analysis is the foregoing second statement B.

For the series-diode circuit of Fig. 8, definitions 1 and 2 apply, where φ_0 and φ_1 are the limits of conduction for $D1$. Again, because of circuit symmetry, we need consider only the half-period $\varphi_0<\varphi<\varphi_0+\pi$. Inductors L are identical with inductor L of Fig. 1. Now, however, it is evident that, to maintain current I, either $D1$ or $D2$ must be conducting, for every φ. It follows that $\theta=\pi$ and that, as a third statement

C. At $\varphi=\varphi_0$ and at $\varphi=\varphi_0+\pi$, there is no voltage drop across the diodes.

The analysis proceeds from this statement and from statement B.

The equations applicable to Figs. 7 and 8 are given in the Appendix, and the resulting functions are plotted in Figs. 9 and 10. For the parallel-diode modulator, the approximately linear range of fractional susceptance is centered at $E_B/E=1$, and it may be shown that $d^2B/dE_B^2=0$ at this point. Similarly, it may be shown that for the series-diode modulator, d^2B/dE_B^2 is nowhere equal to zero. An interesting and possibly useful result for the series-diode circuit is the fact that the fractional susceptance is proportional to the square of the control current over its entire range.

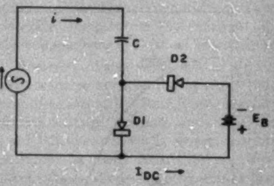


Fig. 7. Parallel-diode reactance modulator

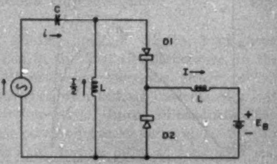


Fig. 8. Series-diode reactance modulator

Effect of Imperfect Diodes

The real performance of diode modulators may be expected to differ somewhat from that described by this analysis; real diodes have appreciable forward resistance, and most voltage sources across which the modulator is connected have appreciable impedance. Because of the difficulty of determining the effect of these impedances quantitatively, no attempt will be made to do so, but a qualitative explanation of the effect of diode forward resistance may be stated simply.

Suppose that the diode in Fig. 1 has zero reverse conductance and a positive forward resistance. For a given current

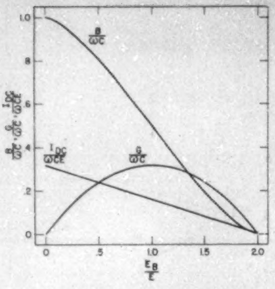


Fig. 9. Variation of susceptance, conductance, and d-c control current, with control voltage, parallel diodes

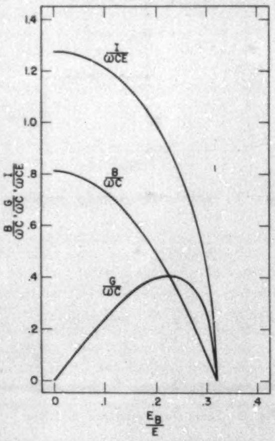


Fig. 10. Variation of susceptance, conductance, and control current, with control voltage, series diodes

i, the diode may be replaced by an ideal diode, as defined previously, in series with a zero-impedance generator of voltage $e_D(\varphi)$. Now,

$$e_D = E_D + e_{D\omega}(\varphi) + \ldots$$

where E_D is the direct component of e_D, $e_{D\omega}$ is the fundamental alternating component, etc. It is evident that E_D aids E_B; to obtain a given current I, therefore, E_B must be smaller than its theoretical value. Voltage $e_{D\omega}$ is in phase with i_ω and may be considered as an additional generator in series with generator e. Its effect is to modify i_ω so that it is more nearly in phase with e. The net result is that, for any I, the fractional susceptance will be less than its theoretical value, and for large I, the fractional conductance will be greater than its theoretical value.

With presently available diodes, however, there will be many applications in which the diode resistance is of no practical importance. Note that the peak diode current never exceeds $2I = 2\omega CE$. If the diode forward voltage drop for a current I is small compared with E, and if the reactive current drawn by the modulator is small compared with the short-circuit current of the source, real performance of the modulator should be close to that predicted by theory.

Experimental Applications

In addition to the previous references, successful use of these modulators has been made in two recent applications. The first[5] uses the circuit of Fig. 1 as an automatic frequency control to correct local oscillator drift in a frequency-modulation receiver. The second[6] uses the circuit of Fig. 7 to frequency modulate a 500-kilocycle oscillator in a wireless microphone. Of particular interest in this case is use of the d-c power $E_B I_{DC}$ to power transistors that amplify the modulating signal.

Appendix

Single-Diode Modulator

Consider the following definitions:

$$e = E \sin \varphi = E \sin \omega t \tag{1}$$

$$\theta = \varphi_1 - \varphi_0 \geqslant 0 \tag{2}$$

Then

$$i = C de/dt = \omega C de/d\varphi \tag{3}$$

for $\varphi_0 < \varphi \leqslant \varphi_1$

and

$$i = -I \tag{4}$$

for $\varphi_1 \leqslant \varphi < \varphi_0 + 2\pi$

From statement A

$$(1/2\pi)\int_{\varphi_0}^{\varphi_0+2\pi} i d\varphi = 0 \tag{5}$$

From statement B

$$\int_{\varphi=\varphi_0}^{\varphi_0+2\pi} ei\, dt = 2\pi E_B I/\omega \tag{6}$$

From equations 3 and 4

$$I/\omega CE = -\cos \varphi_1 \tag{7}$$

From equations 5 and 7

$$\tan \varphi_1 = -\frac{2\pi - \theta + \sin \theta}{1 - \cos \theta} \tag{8}$$

and from equations 6 and 7

$$E_B/E = -(1/4)(\cos \varphi_1) \times (1 - \cos \theta - \sin \theta \tan \varphi_1)^2 \tag{9}$$

which allows the plotting of Fig. 2.

The current of interest is

$$i_\omega = I_Q \cos \varphi + I_R \sin \varphi \tag{10}$$

whose coefficients can be calculated with the Fourier series equations, so that

$$G/\omega C = -(2E_B/E) \cos \varphi_1 \tag{11}$$

and

$$B/\omega C = (1/2\pi)[\theta - \sin \theta \cos \theta + 2 (\cos^2 \varphi_1) \times (2\pi - \theta + \sin \theta \cos \theta + \sin^2 \theta \tan \varphi_1)] \tag{12}$$

Fig. 4 is plotted using equations 7 through 9, 11, and 12.

Parallel-Diode Modulator

The term $\theta = \varphi_1 - \varphi_0 \geqslant 0$ is defined as the angle of conduction for diode $D1$. Then

$$i = C de/dt = \omega C de/d\varphi$$

for $\varphi_0 < \varphi \leqslant \varphi_1 \quad (13)$

and

$$i = 0 \tag{14}$$

for $\varphi_1 \leqslant \varphi < \varphi_0 + \pi$

From statement B

$$\int_{\varphi=\varphi_0}^{\varphi_0+2\pi} ei\, dt = -\int_{\varphi=\varphi_0+\pi}^{\varphi_0+2\pi} E_B i\, dt \tag{15}$$

From equations 13 and 14

$$\varphi_1 = \theta + \varphi_0 = \pi/2 \tag{16}$$

and from equations 15 and 16

$$E_B/E = 1 + \cos \theta \tag{17}$$

The direct-current term is

$$I_{DC}/\omega CE = -(1/2\pi CE)\int_{\varphi=\varphi_0+\pi}^{\varphi_0+2\pi} i\, dt = (1/2\pi)(1 - \cos \theta) \tag{18}$$

From calculation of the Fourier coefficients, as before

$$G/\omega C = (1/\pi) \sin^2 \theta \tag{19}$$

and

$$B/\omega C = (1/\pi)(\theta - \sin \theta \cos \theta) \tag{20}$$

Fig. 9 is plotted with the use of equations 17 through 20.

Series-Diode Modulator

The term $\theta = \varphi_1 - \varphi_0 \geqslant 0$ is defined as the angle of conduction for diode $D1$. From the definition of I, however, $\theta = \pi$. Then

$$i = I/2 \tag{21}$$

for $\varphi_0 < \varphi \leqslant \varphi_0 + \pi$

and

$$i = -I/2 \tag{22}$$

for $\varphi_0 + \pi < \varphi \leqslant \varphi_0 + 2\pi$

From statement C

Vol.
Pt. 1
1958

$$E \sin(\varphi_0+\pi) = E \sin \varphi_0 + \int_{\varphi=\varphi_0}^{\varphi_0+\pi} (I/2C)dt \qquad (23)$$

and

$$I/\omega CE = -(4/\pi) \sin \varphi_0 \qquad (24)$$

From statement B

$$\int_{\varphi=\varphi_0}^{\varphi_0+2\pi} ei\, dt = 2\pi E_B I/\omega \qquad (25)$$

and

$$E_B/E = (1/\pi) \cos \varphi_0 \qquad (26)$$

From calculation of the Fourier coefficients

$$G/\omega C = -(8/\pi^2) \sin \varphi_0 \cos \varphi_0 \qquad (27)$$

and

$$B/\omega C = (8/\pi^2) \sin^2 \varphi_0 \qquad (28)$$

Fig. 10 is plotted with the use of equations 24 and 26 through 28.

References

1. Frequency Modulation of Quartz Crystals by Means of Variable Resistance Devices, Isaac Koga. *Electrotechnical Journal of Japan*, Tokyo, Japan, vol. 4, May 1940, pp. 99–107. Reviewed in "Frequency Modulation of Quartz Crystal Oscillators," *Electronics*, New York, N. Y., vol. 14, Apr. 1941, pp. 98–99.

2. Electronic Frequency Modulator, Gilbert R. Clark. *U. S. Patent no. 2,610,318*, Sept. 9, 1952.

3. Oscillator Frequency Control, Thomas T. N. Bucher. *U. S. Patent no. 2,708,739*, May 17, 1955.

4. The Useful Diode Modulator, Robert H. Weitbrecht. *CQ*, New York, N. Y., vol. 8, Apr. 1952, pp. 18–20.

5. AFC Improves FM Tuners, G. Franklin Montgomery. *Radio-Electronics*, New York, N. Y., vol. 28, Jan. 1957, pp. 123, 126.

6. Wireless Microphone Uses F-M Modulation, G. Franklin Montgomery. *Electronics*, New York. N. Y., vol. 31, Jan. 3, 1958, pp. 54–55.

An Analog Frequency-Measuring Circuit Accurate to 0.1 Per Cent

J. MITCHELL
NONMEMBER AIEE

DURING the course of work on an electromechanical analog computer it was necessary to convert frequencies in the range of 2 to 20 kc into shaft rotations. The shaft rotation had to be proportional to frequency with a departure from linearity of less than ±0.1% of full scale. The circuit was required to operate over the temperature range of −55 to +55 C (degrees centrigrade).

A diode pump circuit,[1] which is a pair of diodes used as a switch alternately to charge and discharge a capacitor, was chosen for its simplicity and consequent reliability. Modifications to the basic circuit were necessary because of the imperfect operation of diodes used as a switch.

This paper discusses the imperfections of diodes as switches and shows how these are compensated to obtain the high degree of linearity and stability.

The Basic Circuit

The diode pump uses each cycle of input frequency to remove a fixed charge from a storage capacitor. The integral of these charges is a current proportional to frequency which is balanced against a current derived from the tap of a linear potentiometer. Any current unbalance is used to control automatically the potentiometer shaft. At balance the position of this shaft is an accurate measure of the input frequency. The basic circuit is shown in Fig. 1. The operation of this circuit at balance may be explained very simply if the diodes D_1 and D_2 are assumed to be perfect switches.

The driving circuit provides currents of about 1 ma (milliampere) which alternately charge and discharge C_1 during each cycle of input signal. The charging current i_1 flows through C_1 and D_1, the discharging current i_2 through C_2, D_2, and C_1. The voltage through which C_1 may swing is limited by diodes D_3 and D_4 to a peak-to-peak value of V_1. Each discharge current takes a charge $Q_1 = C_1V_1$ from C_2. In one second charge, fQ_1 is removed from C_2 where f is the input frequency. This is an average current of fQ_1 amperes. The voltage across C_2 is held at zero by the injection of an equal current through R_1. This current is $i_3 = v/R_1$ where v is the voltage at the tap of the potentiometer.

Hence, at balance,

$$v/R_1 = fC_1V_1$$
$$v = C_1R_1V_1f \qquad (1)$$

In order for v, and therefore the potentiometer shaft rotation, to be proportional to f it is necessary that $C_1R_1V_1$ be constant for all frequencies and temperatures within the specified ranges. C_1 and R_1 are independent of frequency in the audio range and are chosen with equal and opposite temperature coefficients. V_1, the voltage swing of C_1, must be stabilized.

However, diodes which act as perfect switches were assumed during the derivation of equation 1. Real diodes conduct reverse leakage currents and have forward voltage drops. It is necessary to examine the basic circuit to see what effect these diode imperfections may have on its operation.

Diode Imperfections

Silicon junction diodes have been used in the final circuit because, in the OFF condition, they are nearly perfect switches with leakage currents of about 10^{-9} amperes. In the ON condition, they are poor switches requiring about 0.6-volt forward drop in order to conduct 1 ma.

Because of this large forward drop and because of the rapidly increasing diode resistance with reduced forward voltage, the basic circuit will not have the same initial conditions for each charging and discharging current throughout the 10-to-1 range of frequency. Consider only the discharging half of one cycle. When, at the end of a charging cycle, the capacitor C_1 (see Fig. 1) is charged to voltage V_1, V_a will be approximately

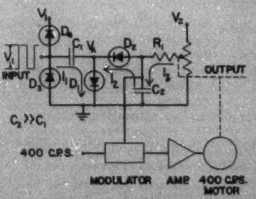


Fig. 1. Basic diode pump

Paper 58-1275, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted December 16, 1957; made available for printing August 28, 1958.

J. Mitchell is with Link Aviation, Inc., Binghamton, N. Y.

The author acknowledges the assistance of N. F. Moody and P. M. Thompson of the Defence Research Board of Canada.

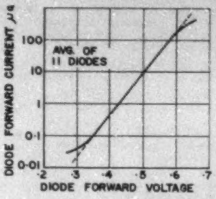


Fig. 2. Silicon diode forward characteristics

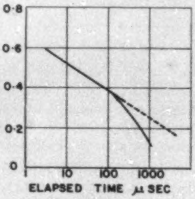


Fig. 3. Decay of V_a

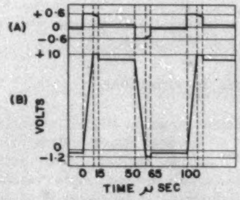


Fig. 4. Required waveforms

A—Pump waveform, V_a
B—Input waveform

+0.6 volt. This voltage will decay at a a rate,

$$dV_a/dT = i_f/C_1 \tag{2}$$

where i_f is the forward current through D_1. Below 100 μa (microamperes), the forward current is given by the expression[2]

$$i_f = i_s(e^{v/CKT} - 1) \tag{3}$$

where

$v = V_a$ the forward voltage
KT = energy equivalent of temperature
C = a constant dependent on diode construction
i_s = theoretical saturation current

Fig. 2[3] shows this current–voltage relation for 11 typical silicon diodes.

From equations 2 and 3 it can be shown that V_a decays logarithmically. This is illustrated in Fig. 3 which is a plot of V_a against log (time). As V_a falls below 0.3 volt surface leakage, not accounted for in equation 3, begins to come into play. Since i_f is greater than equation 3 states, V_a falls more rapidly than is predicted by equations 2 and 3.

The linear relation between v, the potentiometer voltage, and f, the input frequency expressed in equation 1, does not hold because of the slow decay of V_a. This voltage, at the start of a discharge cycle, will be somewhere between +0.5 and +0.2 volt depending on the time which has elapsed since the charging current. The discharging current will therefore remove a charge in the range C_1 $(V_1 - 0.5)$ to C_1 $(V_1 - 0.2)$ from C_2, the exact charge being dependent upon the frequency.

Compensation for Diode Imperfection

The charge removed from C_2 has been made constant by ensuring that V_a returns to a near-zero voltage at a fixed time after each edge. The desired waveform is shown in Fig. 4(A). It is obtained by arranging the limiter so that within 5 microseconds after each charging or discharging the input voltage will fall about 0.6 volt. The necessary waveform is shown in Fig. 4(B). The 0.6-volt step is obtained from the circuit of Fig. 5.

Immediately upon the application of a 1-ma step input current the diode of Fig. 5 conducts almost all the current and V_{in} rises to 0.6 volt. As the series resistance of the choke is low, the current in L rises linearly at the rate V_{in}/L amperes per second. The diode drop changes little until the choke is conducting nine tenths of the current. This occurs in about 4 microseconds afrer which V_{in} falls rapidly to 10 millivolts. A silicon diode is used to form the step in order that its height will vary with temperature in the same way that the drop across the pump diode varies. A 2.7-K resistor is required to damp the ringing which would otherwise occur between the diode capacitance and the choke.

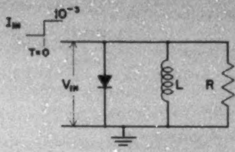


Fig. 5. Step-forming circuit

This circuit is included in the final pump circuit shown in Fig. 6.

Temperature and Supply Voltage Compensation

Two minor additions to the basic circuit ensure that the input voltage swing is largely independent of temperature and that the supply voltage variations will have little effect on the linearity. The first objective is achieved by adding two extra silicon diodes into the reference supply. When this is done, the discharging current reduces the voltage across C_1 by the reference supply V_1 plus the voltage drop across diodes *3*, *4*, and *5* less the drops across diodes *2*, *7*, and *8* as may be seen from the diagram of the input waveform in Fig. 4(B). Thus the voltage swing will vary with temperature only as the difference between diode drops varies with temperature. This is a second-order dependency.

By using the same supply voltage for the reference as for the linear potentiometer, the effect on linearity of variations in this supply is largely eliminated. Only a fraction of any change in supply voltage is reflected as a change in the constant of proportionality of shaft position to frequency. This fraction is the ratio of the difference between diode drops to the total voltage swing and is less than 0.008, and is probably less than 0.004. The supply voltage is stabilized to ±5%.

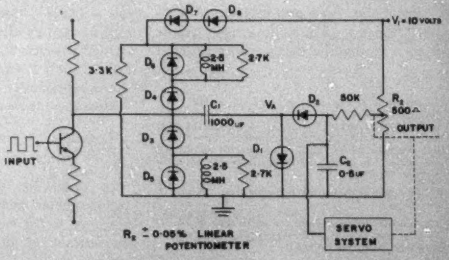


Fig. 6. Final circuit diagram

Detector

In order to measure the range of frequencies at 2–20 kc, C_1 and V_1 were arbitrarily chosen equal to 1,000 micromicrofarads and 10 volts. This makes $R_1=50\,K$ for a full shaft rotation representing 20 kc. Hence the maximum current flowing into the storage capacitor is $V_1/R_1=200$ μa. In order to maintain balance to ±0.1% of full scale, the servo system must operate with a signal of less than ±0.2 μa. The detector used is the modulator described by Moody[3] which can reliably detect signals of ±0.1 μa.

Conclusions

With reference again to equation 1

$v=C_1R_1V_1f$

Now

$v=rV_1$

where $r(<1)$ represents the fractional rotation of the potentiometer shaft. Thus

$r=R_1C_1f$

with the errors noted in the two preceding sections.

The circuit shown in Fig. 6 was built using a linear potentiometer which was calibrated with a potential divider to ±0.01% and tested over a frequency range of 2 to 20 kc as measured on an electronic counter to ±1 cycles per second. The shaft position was linearly proportional to input frequency with an accuracy of ±0.04% of full scale at any temperature and the constant of proportionality changed by 0.1% between −55 C and +55 C.

Summary

By accurately controlling the conditions at the start of each discharging current, by compensating for temperature variations in the critical limiting and current-defining components, and by balancing supply voltage variations, an instrument has been constructed which will automatically convert frequency into shaft position with an accuracy of ±1.0%.

References

1. WAVEFORMS, Britton Chance et al, editors. McGraw-Hill Book Company, Inc., New York, N. Y., "Radiation Laboratory Series," vol. 19, Massachusetts Institute of Technology, 1949.

2. ELECTRONS AND HOLES IN SEMICONDUCTORS, W. Shockley. D. Van Nostrand Company, Inc., Princeton, N. J., 1950.

3. A SILICON DIODE MODULATOR OF 10^{-8} AMPERE SENSITIVITY, N. F. Moody. *Proceedings*, National Electronics Conference, Chicago, Ill., 1955.

◆

Discussion

P. M. Thompson (Defence Research Board, Ottawa, Ont., Canada): Mr. Mitchell shows that it is possible to achieve a linearity of ±0.1% with a diode-pump frequency-measuring circuit if care is taken to compensate it for diode imperfections. In his treatment he considers the leakage current and the forward voltage drop. However, at high switching speeds the effects of diode turn-off time must become important. The turn-off characteristic must bear directly on the fall of voltage of the pump waveform V_a [Fig. 4(A)] during the time between the end of the charging cycle and the collapse of voltage across the 2.5-millihenry inductors shown in Fig. 6. If this is not well controlled, it will be difficult to define the desired potential step due to D_5 and D_6 (Fig. 6) necessary to reduce the potential at V_a to 10 millivolts or less.

J. Mitchell: Mr. Thompson's point about switching speeds is a good one. In fact, different types of diodes may show markedly different characteristics when forward current is plotted against voltage and, therefore, the decay of V_a will vary from type to type. Furthermore, the diode capacitance attenuates the effect of the collapse of voltage across the inductor. I think, however, that the diode turn-off characteristic, in the usual sense of the term, would only indirectly (in that it may imply a greater or lesser effective diode capacitance) affect the circuit's operation.

The maximum frequency which may be measured by this circuit will be determined first by the currents available to charge and discharge C_1 (Fig. 6) and by the value of C_1 itself, and second by the time that V_a must be allowed to decay naturally if the collapse of voltage across the inductor is to be entirely effective in reducing V_a to the same value (±10 millivolts) near zero every time in spite of the diode capacitance. The first mentioned time may, in theory, be made very small. It is doubtful if the second can be reduced below 4 to 5 microseconds. Therefore, there is an absolute frequency limit of about 100 kc. The practical frequency limit is undoubtedly lower, probably no greater than 40 to 50 kc.

Organic-Moderated Reactors for Central Station Power

W. E. PARKINS
NONMEMBER AIEE

E. F. WEISNER
NONMEMBER AIEE

THE CONCEPT of a power reactor which would use an organic fluid as the moderator and coolant has been considered an attractive possibility for several years. This concept suggested itself because of the many physical properties of organic fluids which are superior to those of water when used as the moderator in a nuclear reactor.

In its simplest form an organic-moderated reactor would utilize a tank containing the core and through which the organic fluid would be pumped to remove the heat being generated. The fuel would be in the form of an array of solid elements within the core tank, and appropriate control elements would be distributed for startup, shutdown, and regulation of the reactor power.

As in the case of water-moderated reactors, the fuel could not be natural uranium but would of necessity contain some form of enrichment. The nuclear properties as they affect the reactor performance are very similar for water and for organic fluids of interest. The principal difference between a water-moderated reactor and an organic-moderated reactor in the simple description just given, would be in the design of the core tank. The organic fluid, being of lower vapor pressure, would be operated at the same or higher temperatures, while using a vessel designed for about one tenth the pressure which would be required for a water-moderated system.

Choice of Organic Fluid

In choosing a suitable organic fluid to serve as the moderator-coolant in a power reactor, it was necessary to investigate the stability of many different hydrocarbon compounds. These

---

Paper 58-1272, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 16, 1958; made available for printing September 9, 1958.

W. E. PARKINS and E. F. WEISNER are with Atomics International, Division of North American Aviation, Inc., Canoga Park, Calif.

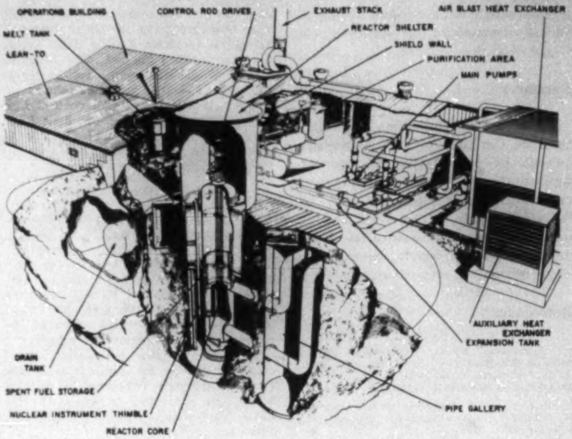


Fig. 1. OMRE facility at the National Reactor Testing Station

investigations included both stability against thermal, or pyrolytic, decomposition and stability against radiation decomposition. The results of many measurements carried out at elevated temperatures in the laboratory and in special exposures with gamma-ray, electron, and neutron radiation have indicated the polyphenyl compounds to be the most suitable class. These compounds are formed by linked aromatic rings. Of greatest interest are diphenyl, consisting of two rings, and the ortho-, meta- and para-isomers of terphenyl, based on the different linkages of three aromatic rings.

The diphenyl and commercial mixtures of the terphenyl isomers are available at prices of less than 20 cents per pound. During operation of the reactor, it is necessary to add new organic material because, while these compounds are quite stable, they do gradually decompose under the influence of the radiation present in the reactor core.

Advantages of an Organic Fluid as Reactor Moderator

Because of the lower vapor pressures, organic fluids may be used in the 500–700 F (degrees Fahrenheit) range without the requirement of a high-pressure system. Diphenyl boils at 492 F at atmospheric pressure, and its vapor pressure at 700 F is 112 psia (pounds per square inch absolute). The terphenyls boil in the neighborhood of 700 F at atmospheric pressure. Practical operating pressures for organic reactors appear to be in the 50- to 300-psia range.

Another paramount advantage for organic fluids is the low corrosion rate. Standard engineering materials such as low-carbon steels and aluminum may be used throughout the reactor and external coolant system, since their corrosion rate with organic fluids is essentially zero. The corrosion rate with uranium and other nuclear fuels is very low, which is important since it means that few radioactive fission products would get into the coolant stream even if the fuel cladding should fail. From the safety standpoint it is to be noted that no exothermic chemical reactions can occur under any possible conditions between the organic fluid and any materials used in the reactor system.

The induced radioactivity in hydrogen and carbon as a result of exposure to the neutrons in the reactor core is negligible. This is important since it results in minor shielding requirements for the external coolant system and greatly eases maintenance problems. In this respect organic fluid is superior to water, which contains oxygen and requires special shielding of the external system during reactor operation.

Problems Associated with Organic Fluids

The principal problem associated with the use of organic fluids in nuclear reactors is the radiation decomposition. The interaction between the nuclear radiation and the organic, results in the production of hydrogen and light hydrocarbons which appear as gases and must be vented. In addition, higher molecular weight or polymeric materials are formed and must be gradually removed by a purification system to maintain their concentration at some suitable equilibrium value. As will be mentioned later, no difficulty has been experienced from the fouling of heat transfer surfaces by the deposition of these high polymers. In the operating range of interest they are maintained in solution in the organic fluid. Similarly, in the operating range of interest there has been no significant contribution to the organic decomposition rate due to pyrolytic effects.

Although most organic materials are flammable, the polyphenyls are not particularly difficult to handle. They have rather high spontaneous ignition temperatures in the neighborhood of 1,000 F. Ordinary refinery practices and precautions are adequate to insure against fires. Because of their low vapor pressure, handling of the polyphenyls at ambient temperatures presents no particular health hazard. Proper ventilation must be provided in locations where it is necessary to work with these fluids at elevated temperatures.

Another disadvantage of these organics is the fact that their melting points are somewhat above room temperature. Diphenyl melts at 157 F, while the commercial mixtures of terphenyls melt in the range of 150–300 F. As a result, it is necessary to provide steam tracing or some form of electric heating for the piping and other components of the organic system.

Design of the OMRE

In order to test the use of an organic fluid used as the moderator-coolant in a reactor, the Atomics International Division of North American Aviation entered into a contract with the Atomic Energy Commission in mid-1955 for the construction and operation of the organic-moderated reactor experiment (OMRE). While many aspects of the organic fluid as used in a reactor were to be investigated, the principal factor to be studied was the organic decomposition and its effect on the over-all reactor performance. As a result, the OMRE was designed for a variety of operating conditions up to a maximum thermal power of 16 mw (megawatts). No provision was made for the generation of electric

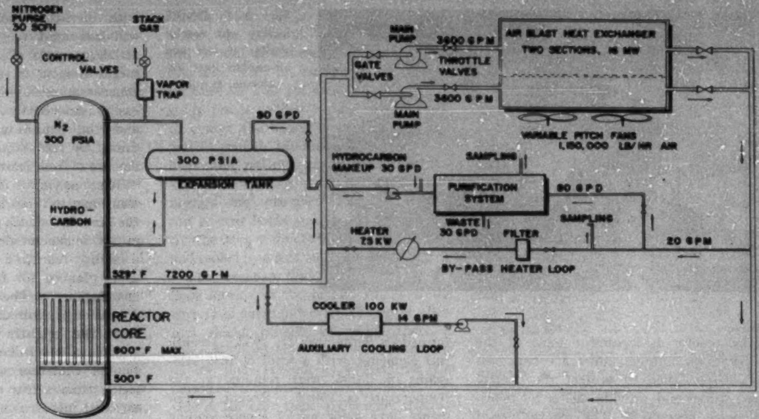


Fig. 2. OMRE schematic flow diagram

power. The reactor heat was dissipated to atmosphere in air-cooled exchangers.

The OMRE was constructed at the National Reactor Testing Station near Idaho Falls, Idaho. Installation was completed in May 1957, and the experimental program has been in progress since that date. The following sections give a brief description of the OMRE design.

Reactor

Fig. 1 shows a cutaway view of the OMRE facility. The reactor vessel is of a mild alloy steel, $4^1/_2$ feet in diameter, 28 feet tall, with a wall thickness of 1 inch. The reactor core is located in the lower region of this vessel.

Because of its availability a plate-type fuel element with fully enriched uranium in the form of UO_2 in a stainless-steel matrix and clad by stainless steel was used. The UO_2 and stainless-steel matrix is 0.020 inch thick and the cladding 0.005 inch thick. The spacing between plates for passage of the organic fluid is 0.134 inch. Each element is assembled from 16 fuel plates enclosed in a rectangular box 2.8 by 2.9 inches in cross section. The uranium fuel is loaded into each fuel plate over an area 36 inches high by 2.5 inches wide. The rectangular boxes comprising the elements are positioned on a 4.5-inch-square lattice and are supported by an upper and a lower grid plate. Thirty-six fuel-element positions are available and an extra position is reserved for an antimony-beryllium neutron source.

A 2-inch-thick thermal shield is located inside the reactor vessel and outside of the fuel-element region. The organic coolant enters the reactor vessel at a plenum below the core. It then flows upward between the fuel plates and discharges into the space between the fuel-element boxes. After flowing downward around the outside of the fuel elements, it flows upward between the thermal shield and the reactor tank and discharges into a 14-foot-deep pool above the reactor core.

The OMRE is controlled by 12 control rods symmetrically located between fuel elements. These rods consist of $1^1/_4$-inch-diameter steel tubes loaded over a 36-inch length with compacted boron-carbide powder. The 12 rods are hung in pairs from six hanger rods and are driven through rack-and-pinion drives. All rods perform the double function of control and safety, with one pair being driven through a differential gear to provide both fast regulating action and slow shim action. For rapid shutdown the rods are released by a ball-type latch, and a kick-off spring insures downward acceleration.

Process System

A schematic flow diagram for the OMRE is shown in Fig. 2. Nitrogen cover gas at a nominal operating pressure of 200 psia is used over the free organic surface in the reactor vessel. A continuous purge of the nitrogen removes the hydrogen and light hydrocarbon gases formed by the organic decomposition. These gases are diluted by the building ventilation system and vented out the exhaust stack.

The cooling system circulates 7,200 gpm (gallons per minute) of organic fluid from the reactor through the air-cooled heat exchanger by means of two centrifugal hot-oil pumps operating in parallel. A small auxiliary coolant loop rated at 100 kw is provided to remove afterglow heat from the reactor during shutdown, when the main coolant loop is inoperative. The auxiliary loop uses a spray cooler as its heat sink. A batch-type purification system is incorporated to remove the high polymers formed in the organic fluid. A still is used to separate the polymer, with the undamaged organic being returned to the primary circulation system together with the necessary makeup material through the expansion tank. This expansion tank is necessary to provide for the change in volume which occurs in the organic during changes in system temperature. In order to measure the heat transfer properties of the organic fluid being circulated in the primary system, a special by-pass loop is incorporated. This loop contains a heater and appropriate means for temperature and flow measurement to permit determination of the heat transfer coefficient.

With the exception of the reactor vessel, the entire organic system is made from carbon steel. All components and piping can be preheated to a temperature of 350 F. Sixty-cycle induction heating is used on the vessels and piping. The pumps and valves are traced with heating cable, and the air-blast heat exchanger is provided with an oil burner.

Plant Layout

The OMRE was constructed as a minimum cost facility. The general layout is shown in Fig. 1 and the completed installation is shown in Fig. 3.

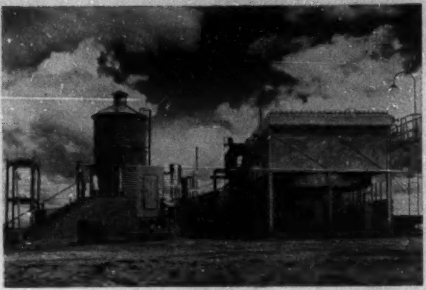


Fig. 3. OMRE facility with reactor shelter at left and air-cooled heat exchanger at right

The reactor is located outside of the operations building under a corrugated-sheet-steel shed to protect the control rod drives and provide shelter for maintenance personnel. The reactor vessel is for the most part buried underground. Inlet and outlet coolant pipes pass through tunnels and to ground level at the operations building. The pumps, purification system, and other auxiliary equipment are located in the process area. The air-cooled and auxiliary heat exchangers are located out-of-doors.

The process area is separated from the control room by a 2-foot-thick wall of ordinary concrete. Valve mechanical operators pass through this wall. The control room houses the operator's console, recording instrumentation, and electric switchgear. Adjacent to it is a separate service area containing the nitrogen-gas supply system and the organic melting and filling system.

Results of the OMRE Operation

As part of the preoperational testing, many portions of the OMRE installation were separately checked. The 60-cycle induction-heating equipment for the piping primary system was operated and found to be very satisfactory. Kerosene was first used in place of the organic to prove out the circulation system. The response of the control-safety rod drives was determined to be satisfactory. The methods for fuel handling were checked out, as well as the operation of the purification and other auxiliary systems.

The organic fluid chosen for first use in the OMRE was Santowax-OM, a commercial mixture principally containing the ortho- and meta-terphenyls together with some diphenyl. After the reactor and primary system were filled with this material and the circulation tests completed, the critical experiment was performed on September 17, 1957. Twenty fuel elements, with a total of 16.4 kilograms of U-235, were required at the temperature of 250 F. Recent experiments have made use of a core loading of a total of 31 fuel elements. Some of these elements are instrumented with thermocouples to provide temperature measurements during operation.

Experiments were carried out to calibrate the control rods, to measure neutron flux distributions within the core, and to determine the various coefficients of reactivity. The reactor performed well and there was no significant unexpected behavior.

Special attention was given to measurement of radiation levels. Trace amounts of inorganic impurities were determined to be present in the coolant but their contribution to the radiation levels was rapidly decreased as a result of operation of the purification system. These impurities, which include sodium, manganese, chlorine, and copper were largely removed together with rust, metal filings, welding slag, and other impurities present within the primary system. Actual measurements of radiation levels at various points during power operation have confirmed the most optimistic predictions. The highest radiation level in the process area was 80 milliroentgens per hour at the surface of the 10-inch primary pipes, while operating at a power level of 10 mw. The radiation level in the center of the process area under these conditions was 13 milliroentgens per hour. Such levels permit access by maintenance personnel on an almost continuous basis.

The principal matter for investigation has been the rate of polymer formation as a function of various operating conditions. The equilibrium polymer content in the circulating organic has been varied up to 30%. The rate of polymer formation was determined to decrease with increasing polymer concentration and has most recently been measured at $2^{1}/_{4}$ pounds of polymer formed per megawatt-hour of operation at the 30% concentration. This material has been easily removed by the purification still and the equilibrium condition maintained by the addition of an equivalent amount of fresh Santowax-OM.

The polymer formation rate just mentioned is known to be about twice the amount which would be obtained in an organic-moderated reactor using a slightly enriched uranium fuel, such as is planned for future power reactor installations. The reasons for this difference is primarily due to the fact that the slightly enriched uranium absorbs a much larger fraction of the gamma radiation present in the core region than is true of the special highly enriched fuel elements being used in the OMRE. All estimates have shown that with current prices for the organic, the contribution of the cost of makeup material to the final cost of electric power produced is less than 1 mill per kw-hr (kilowatt-hour), based on the experimental results which have now been obtained.

As another extremely important matter related to the polymer formation in the organic, no evidence of surface fouling has been observed in the OMRE. Such an effect would have been detected by the fuel-element thermocouples or the by-pass heater loop. These results verify earlier measurements of various organic fluids circulated in in-pile loops at the materials testing reactor.

The operation of the OMRE has shown the organic-moderated type of reactor to be a practical concept. No unusual difficulties have been encountered. The low radiation levels have verified the feasibility of contact maintenance during operation on all parts of the system except the reactor core. A simple type of purification system has been shown to be adequate, and the rates of polymer formation and gas evolution are nominal. The lack of any surface fouling effect has been of special significance in assuring continued high performance of the reactor system.

OMR for the City of Piqua

At the time of the request by the Atomic Energy Commission for proposals for the second round of demonstration reactors, the city of Piqua, Ohio, proposed the construction of an organic-moderated reactor (OMR) for its municipal system. A contract between the Atomic Energy

Vol.
Pt. 1
195

Commission and Atomics International for the engineering and construction of this plant has recently been completed. Preliminary design is in progress and a schedule has been established leading to completion of construction during 1961.

The necessary turbine equipment is available at the Piqua site, and it is required that the reactor system provide steam at the rate of 162,000 pounds per hour, at 450 psia and 550 F steam conditions. The reactor installation is essentially sized for a plant of 12.5 gross electrical mw, or 11.4 net electrical mw.

In addition to the work being carried out on the OMRE, there has been a research and development program sponsored by the Atomic Energy Commission to investigate the properties of organic fluids and to develop components suitable for organic-moderated reactor systems. Additional development work will be carried out specifically in connection with the plant being designed for the city of Piqua. In the paragraphs which follow, a description is given of a reactor of the size required for Piqua. The features described are essentially those now planned as a part of the preliminary design in preparation.

REACTOR

Fig. 4 shows a vertical section of the reactor. Certain of the features are similar to those incorporated in the OMRE. The reactor vessel is of low-carbon steel, has an inside diameter of 6 feet, an over-all height of 25 feet, and a wall thickness of 1 inch. It is designed for a maximum operating pressure of 300 psi gage at 650 F. A guide pin is welded to the bottom head to center the vessel during installation and to provide lateral support. Coolant inlet and outlet nozzles penetrate the vessel above the reactor core. Other nozzles are provided for the control rod electric leads and for thermocouples. At the top, a reinforced head is bolted and sealed by a soft metal gasket.

The reactor tank is located below floor level. In the neighborhood of the core it is surrounded by the outer thermal shield, outside of which there is thermal insulation and a carbon-steel cavity liner serving as a form for the concrete foundation. Immediately inside of the core tank is the inner thermal shield which is supported on a flange welded to the vessel at an elevation above the core. Attached to this inner thermal shield are the upper and lower grid plates, each approximately 6 inches thick. These gridplates are pierced by 69 holes appropriately to support and locate the fuel elements.

The organic fluid planned as the moderator-coolant in the Piqua OMR is Santowax-R. This material consists of a commercial mixture of the three terphenyl isomers. It melts in the range of 140 to 290 F and boils at atmospheric pressure at a temperature of about 700 F. It serves not only as the moderator and coolant, but as the reflector in the region surrounding the core. In addition, a 15-foot-deep pool of organic above the core provides shielding in that region. The organic enters the reactor vessel above the core, then flows downward between the inner thermal shield and the core tank wall. From the lower plenum it flows upward through the fuel-element assemblies as well as through the spaces between fuel elements into the upper pool. Nominal temperature conditions for the organic fluid are 550 F inlet and 600 F outlet. The total flow rate is 11,000 gpm.

FUEL ELEMENT

The reactor fuel is slightly enriched metallic uranium with stabilizing alloying. Required enrichment is 1.8% of U-235 in the uranium. Each fuel element consists of finned aluminum-clad uranium plates contained within a square steel box, as shown in Fig. 5. The fins are of triangular cross section and run parallel to the direction of coolant flow. In addition there are fins of rectangular cross section which extend across the coolant gap to maintain the spacing between fuel plates. The fuel elements are supported in the lower grid plate and are free to expand upward through the locating hole in the upper grid plate. The fuel element boxes are approximately 5.25 inches square and are located on a square array 6 inches between centers. The length of the uranium-loaded section in the fuel plates is 4.5 feet. An orifice at the lower end

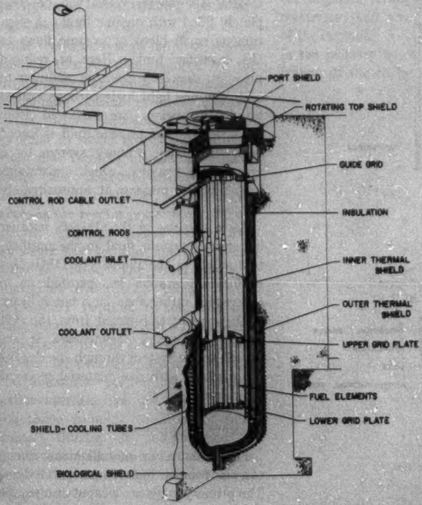


Fig. 4 (left). Reactor for 11.4-mw plant

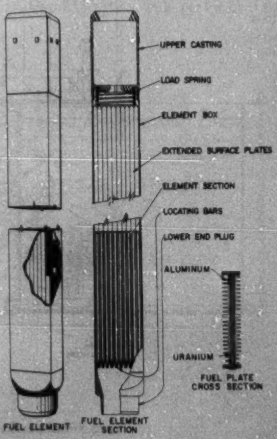


Fig. 5 (right). OMR fuel element

of each fuel element is used to regulate the flow so that the maximum surface temperature within that element does not exceed 750 F.

Control-Safety Elements

Six control elements regulate the reactor and provide for emergency shutdown. The poison rod, drive, and position indicator of each control element are arranged in a compact assembly that can be inserted into a standard fuel-element position in the core. These elements are of the "magnetic jack" type in which movement of the rod is achieved by alternate energizing of a series of magnets to hold, lower, or raise the poison rod in small step increments. The movable portions are sealed within a stainless-steel tube and the only outside connection is by means of the electric leads. The poison rod is released to fall by gravity into the core region whenever the current supply to the hold magnets is interrupted.

Fuel-Handling System

The principal components of the fuel-handling system include the reactor top head, the top shield, the fuel cask, the cask carriage, and the storage areas for new and irradiated fuel. The general arrangement is shown in Fig. 6.

Access to the core tank is provided by loosening the flange bolts which fasten the upper head, and raising this head slightly so that it may be attached to and rotated with the top shield. The shield is supported on a circular bearing which allows rotation to any desired angle or position. The top shield assembly contains a plug which fills a port in the shape of a radial slot in the top shield. By properly rotating the shield, the access port may be brought over any particular position within the core.

A lead-shielded cask mounted on a carriage is then moved to the top shield and located over the desired position. Fuel elements may be withdrawn into the cask by means of a latching assembly which extends down through the cask. The escape of organic vapors during this period is minimized by a sliding assembly over the port of the top shield. Fuel-changing operations would normally be carried out with the organic temperature in the neighborhood of 350 F. It should be noted that the organic fluid is quite opaque and these manipulations must be carried out without the aid of visual observation in the core tank.

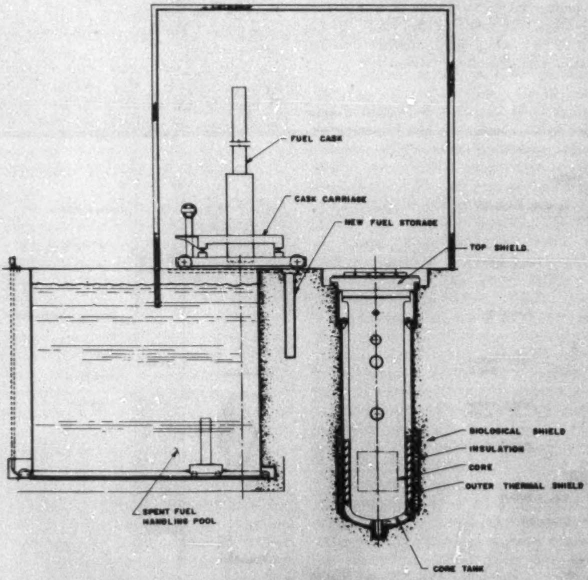


Fig. 6. Fuel-handling arrangement for OMR

After an aligned element has been withdrawn into the cask, the carriage can be traversed over an adjacent pool of water used as a storage facility. Fuel elements may be lowered into this canal and handled in a rack arranged to move along its floor. After a period of storage for the decay of afterglow heating, the fuel elements may be loaded into a cask for shipment to a chemical processing plant.

Primary Coolant System

The primary coolant system is divided into two parallel loops. A simplified flow diagram showing one of these loops is given in Fig. 7. From the reactor the coolant flows through a pump, a superheated, and an evaporator, and returns to the reactor. Each pump is a centrifugal unit of the horizontal split-case type with mechanical seals. The superheater and evaporator are shell and tube units horizontally mounted and constructed of mild ssteel. Blocking valves, not shown in Fig. 7, may be used to isolate either loop near the reactor. These valves and others in the system are of the cast-steel back-seated packed-stem type. Coolant lines are seamless carbon-steel pipe. Equipment such as valves, pumps, and other components which require maintenance are flanged into the system with seals made of soft iron gaskets.

In contrast to the arrangement on the OMRE, the reactor vessel will be completely filled with organic fluid. A continuous small bleed of organic flows to the degasifier tank. Here at reduced pressure, dissolved gases, in addition to hydrogen and light hydrocarbons formed by decomposition, are removed from the system. The organic fluid is forced back into the primary system by a pressurizing pump which maintains proper system pressure at approximately 120 psia.

As in the case of the OMRE, another line carries organic fluid to the purification still. In the Piqua-type OMR this purification system is operated on a continuous rather than on a batch basis. High polymer is rejected from the still bottoms. Organic condensate is returned to the system through the degasifier together with new material required for makeup.

Plant Arrangement

Fig. 8 shows an over-all arrangement of a plant similar to the Piqua installation. The primary loops are located in separate

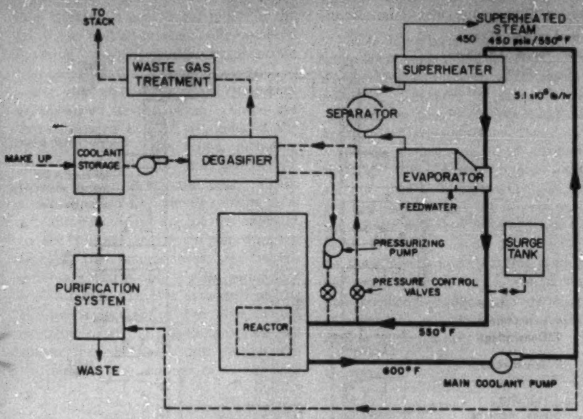


Fig. 7. OMR basic flow diagram

rooms shielded with approximately 2 feet of ordinary concrete. Suitable ventilation control is provided these areas. Steam at 450 psia and 550 F is carried from the superheaters to the turbine located in an adjacent room. A regenerative steam cycle incorporating three stages of steam extraction for feedwater heat is employed. There is a deaerator in the 40-psi extraction line. A conventional a-c alternator is used to generate the 12,500 kw of gross electric power. Feedwater is returned to the evaporator at 337 F. Dry steam leaves the separator before entering the superheater at 448 F.

The control system adjusts reactor power output to that required by the turbine generator while simultaneously maintaining steam pressure constant at the turbine throttle. A signal proportional to steam flow is taken from a steam flow recorder on the main steam header and inserted directly into a neutron flux controller. This signal serves as the reactor power level primary control. The same signal is taken from the steam flow recorder and routed to a power computer. The power computer determines the reactor coolant outlet temperature required to maintain constant steam pressure and sends a signal proportional to this desired temperature to the temperature-recorder-controller. The automatic portion of the control system operates over the power generation range from 15 to 110% of rated power. Startup is accomplished by manual manipulation of the control elements using the nuclear startup instrumentation.

OPERATIONAL AND ECONOMIC ASPECTS

As has been apparent from the foregoing discussion, no steel containment vessel is planned for housing the reactor installation. Characteristics inherent in the design of an OMR make it unnecessary to provide special housing of this nature to prevent the dispersal of radioactivity as a result of any conceivable accident. The principal factors contributing to this inherent safety are the absence of potential chemical reactions, the negative temperature coefficients of reactivity, and the low vapor pressure of the coolant. A sudden loss of pressure in the primary system does not produce boiling of the coolant, and it is possible to design the system to maintain core cooling in spite of any accidental condition.

As was pointed out for the OMRE, it is expected that maintenance of the OMR plant will be straightforward. All components, with the exception of the core, should be accessible. It has been calculated that even a fuel-cladding break will result in small increase in the radiation levels. Again, the purification and degasification systems serve to remove such accidental fission product contamination on a continuous basis.

The capital charges for the second or subsequent complete 11.4 net electrical mw OMR plants are estimated to be approximately $500 per net installed kw or $5,700,000 total. Treating the capital investment as an annual charge of 15% of the total, and assuming an 80% load factor, the capital investment contributes 9.5 mils per kw-hr to the total power cost.

Annual fuel costs depend to some extent on the fuel exposure time and on the load factor. In this reactor the average exposure time is expected to be from 3,000 to 4,000 megowatt-days per per ton. The load factor is estimated at from 60–80% depending upon the application. Within these ranges the fuel cost will be between 5.4 and 6.3 mils per kw-hr. This figure includes the burned-fuel cost, fuel fabrication, processing, shipping, inventory costs, and credit for plutonium produced.

At this stage of development the exact cost for plant maintenance is difficult to predict. Based upon information available it is estimated that the costs for operation, maintenance, and organic makeup will total 2 to 3 mills per kw-hr.

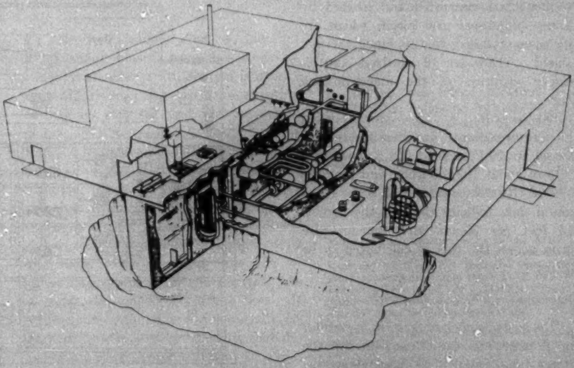


Fig. 8. Cutaway view of 11.4-mw OMR plant

The total cost of power from this plant is therefore in the neighborhood of 18 mills per kw-hr. Reductions in this figure are expected as further advances are made in the technology of organic-moderated reactor plants.

OMR for Large Central Station Plant

The organic-moderated reactor concept, while very suitable for small, inexpensive, power-reactor installations, is not limited to plants of small output. Engineering studies and preliminary designs have been carried out for plants of 50, 100, and 150 net electrical mw output.

These designs have all been based on the use of a single reactor in each plant, and no limitation has yet been foreseen in the maximum thermal power which might be obtained from a single reactor installation. The lack of such a limitation in the maximum output is in part due to the fact that the primary system, and hence reactor vessel, operates at a relatively low pressure.

To give an example of a large OMR central station plant, the following paragraphs describe some of the features recommended for a plant of 150 electrical mw.

REACTOR

A vertical cross section of the reactor is shown in Fig. 9. As is apparent, except for size the reactor design is very similar to that described for the Piqua-type OMR. The reactor vessel is approximately $14^1/_2$ feet in diameter and 49 feet high. The active core region is 10 feet in height, 11.3 feet in diameter, and contains 325 fuel elements. These elements are supported and located by lower and upper grid plates which are in turn attached to the inner thermal shield. There are 16 unitized control-safety elements of the magnetic-jack type.

Santowax-R is the recommended moderator-coolant. It enters the reactor vessel at 550 F, exits at 618 F, flows at a rate of about 100,000 gpm, and removes a total thermal power of 555 mw. Maximum flow velocity through the core is 15 feet per second and the pressure drop across the core is calculated to be 35 psi.

PROCESS AND STEAM SYSTEMS

The organic-coolant system is similar in principle to that described for the Piqua-type OMR. It uses, however, six parallel heat-removal loops. Each

Fig. 9 (right). Vertical section of reactor for a 150-mw plant

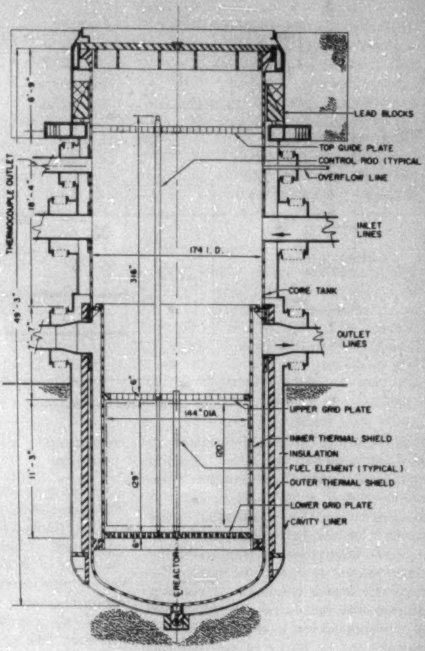
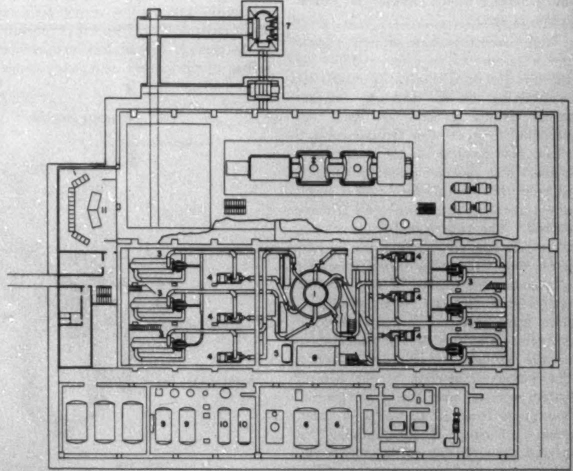

6'-9"
LEAD BLOCKS
TOP GUIDE PLATE
CONTROL ROD (TYPICAL
OVERFLOW LINE
THERMOCOUPLE OUTLET
18'-4"
INLET LINES
174 I. D.
316"
CORE TANK
49'-3"
7'-7"
OUTLET LINES
6"
144" DIA
UPPER GRID PLATE
120"
INNER THERMAL SHIELD
INSULATION
FUEL ELEMENT (TYPICAL)
OUTER THERMAL SHIELD
11'-3"
129"
LOWER GRID PLATE
CAVITY LINER
6"
℄ REACTOR

Fig. 10 (below). Plan arrangement of 150-mw plant

1—Nuclear reactor
2—Turbine-generator unit
3—Steam generator and superheater
4—Primary coolant pumps
5—Coolant surge tank
6—Reactor drain tanks
7—Main power transformer
8—Spent-fuel storage pit
9—Degasifier
10—Waste receiver tank
11—Control room


7
11
3
4
1
5
6
9
10

loop contains a centrifugal pump in the hot leg as well as automatically operated blocking valves capable of isolating that loop from the others.

Fig. 10 is a plan view of the equipment arrangement in the plant. There are three loops on each of two sides of the reactor. These three loops, including their evaporators and superheaters, are housed in a single process room. Tanks and equipment for organic-coolant storage and purification are located adjacent to the process area. On the opposite side of the process area, the turbogenerator and steam system equipment are installed. In this arrangement the turbine is located on the floor at the elevation of the top of the reactor. An extension of this floor forms the concrete shielding at the ceilings of the process areas.

A total of 320,200 pounds of steam per hour are produced at 415 psia, 550 F. The turbine is a high-speed unit having a gross capability of 162,000 kw and using $1^1/_2$ inches of mercury exhaust pressure. Three feedwater extraction stages are provided.

OPERATIONAL AND ECONOMIC ASPECTS

The building housing the OMR installation is of conventional construction and no special containment vessel need be provided. The inherent safety features described for the Piqua-type OMR also apply to this larger plant.

In regard to costs, it is estimated that capital investment required for the 150-mw OMR is approximately $250 per net installed kw. Fuel costs on the same basis described for the Piqua-type OMR are 4–5 mills per kw-hr. Based on these figures and including appropriate charges for operation, maintenance, and terphenyl makeup, the total cost of power is estimated to be approximately 10–11 mills per kw-hr.

Summary

The basic experiments on the properties of organic fluids and the recent operation of the OMRE have shown the concept of a nuclear reactor moderated and cooled by an organic fluid to be practical. This type of reactor appears to have particular advantages for central station application because of the adequate steam conditions obtainable and the excellent safety characteristics of OMR plants. Capital costs are low because of the nominal operating pressures required and because conventional materials, together with conventional field fabrication techniques, can be used throughout the installation.

Work on design of the 11.4 net electrical mw OMR for the City of Piqua, Ohio, is in progress. New information is currently being obtained as part of the research and development being carried out for the organic reactor program.

The concepts developed for the Piqua-type OMR can be applied to much larger central station plants. There appears to be no basic limitation in the size of such plants.

Present estimates indicate the cost of power to be comparable to, if not lower than, that obtainable from other types of nuclear plants. Actual performance of large OMR plants, as well as of other types, must await the additional experience which will follow when such plants are built and operated.

Theory of Transient Build-up in Avalanche Transistors

W. SHOCKLEY
NONMEMBER AIEE

J. GIBBONS
NONMEMBER AIEE

Synopsis: It is found that an avalanche transistor biased into the negative resistance range will deliver an exponentially increasing current into a capacitor connected between the emitter and collector. For a uniform base layer of width W, minority carrier diffusion constant D, emitter efficiency γ, collector capacitor C_c, load capacitor C, and avalanche multiplication M, the current builds up as $\exp \lambda t$ where

$$\lambda=(D/W^2)[\cosh^{-1} \gamma MC/(C+C_c)]^2-\nu$$

where ν is the reciprocal of the minority carrier lifetime. The time required for a voltage drop from near breakdown voltage to about 89% of breakdown voltage is about $W^2/4D$ provided $C_c/C<<1$ and γ is nearly unity; the rate of drop of voltage is substantially equal to $2V_BD/(n+1)W^2$ after the voltage has dropped from near V_B to below 90% V_B where n is the exponent in $1/M=1-(V_c/V_B)^n$. If the ratio of C/C_c is too small, the collector voltage will not "bottom," i.e., drop to the base voltage; means of estimating the critical size of the capacitor are given. The analysis is based on an integral equation involving stored minority carrier charge in the base layer, which may be useful in other transient problems of transistors.

IT IS well known that a junction transistor biased into the range at which avalanche multiplication is important will discharge a capacitor connecting emitter and collector, and generate a wave of voltage which has high harmonic content. The analysis of this discharge wave poses difficult mathematical problems, although the physics of the problem, in principle at least, is relatively simple.

The transition results from feedback due to secondary carriers arriving back at the base, and provoking injection of minority carriers from the emitter. As a result, the collector current rises. It will be shown that this rise is initially exponential, with a time constant that depends on the collector voltage.

The exponential increase continues until the collector voltage drops appreciably. From this point on the problem becomes mathematically more difficult, because it is essentially nonlinear in the voltage and current variables.

The problem is treated in the large signal range by making use of the concept of the stored minority carrier charge in the base layer. This quantity, denoted by Q_s, initially increases exponentially but approaches a maximum value as the avalanche multiplication drops. The variation of Q_s with time can be analyzed with the aid of an integral equation which can be treated approximately.

The general conclusions reached are as follows:

1. During the initial constant voltage phase of the build-up, the collector current I_c and the stored charge Q_s build up exponentially as $\exp(\lambda t)$ where

$$\lambda=(D/W^2)(\ln 2M)^2$$

where D is the diffusion constant of minority

Paper 58-1249, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 10, 1958; made available for printing September 9, 1958.

W. SHOCKLEY and J. GIBBONS are with the Shockley Semiconductor Laboratory, Mountain View, Calif.

The work reported in this paper was supported in part by a contract with the U. S. Signal Corps.

carriers in the base layer, W is the width of the base layer, and M is the avalanche multiplication factor for minority carriers entering the collector junction from the base. For $M=10$, the squared factor is 9, and for $M=100$, it is 28. Thus multiplication can greatly increase the rate of build-up of current.

2. The larger the load capacitor, the faster the drop in voltage across the capacitor. This result is at first surprising. It is derived assuming that a linear approximation may be used in treating minority carrier motion in the base layer but may be more generally valid. The external capacitor can be considered as large when it has many times the capacitance of the collector junction. If the load capacitor is not large compared to the collector-junction capacitance, the transistor may not be able to bottom, i.e., completely discharge the load capacitor with a sharp voltage wave.

3. The current into the load capacitor rises quite abruptly and reaches nearly its peak value before the capacitor is substantially discharged. The time required for the current to increase from 25% to 75% of its peak value is of the order of $W^2/4D$.

The results are based upon an especially simple process for turning on the transistor. As discussed in the following section, it is possible to consider a situation in which the transistor is biased by high-impedance current sources so that it is in a negative-resistance condition between emitter and collector terminals. If a capacitor, biased to the same voltage, is connected across these terminals, an unstable situation is produced such that an initial transient of increasing current will build up.

Important practical cases correspond to somewhat different initial conditions. For example, an effective varying alpha may result from connecting a resistor between base and emitter; then the transient may be initiated by increasing forward bias across the emitter junction. Alternatively, a voltage pulse may be applied to drive the collector momentarily into the breakdown range. Such cases will require treatments differing from that presented here, but many of the essential features will be the same.

The Constant-Voltage Case

In Fig. 1 the sign conventions used in this treatment are illustrated. The currents I_e, I_b, and I_c are the total currents flowing into the terminals of the *p-n-p* transistor. The current I is introduced to eliminate minus signs in the equations that follow.

Consider first the situation corresponding to the switch S being open. The resistor, supposed to be very large in value, charges the load capacitor C to the potential of the collector terminal. The battery of potential $-V_B$ is discussed in the next section. Closing the switch S changes the impedance across the emitter to collector terminals and leads to the transient build-up discussed in the following.

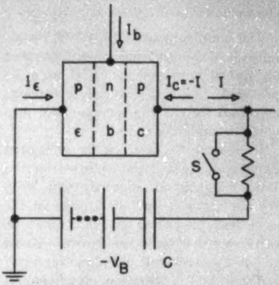


Fig. 1. Circuit for establishing initial conditions for small transients

First, several quantities concerned with holes will be defined:

γ = emitter efficiency = hole current injected across emitter junction divided by I_e

and the base-layer transmission factor β

$$I_{mc}=\beta\gamma I_e+I_{co}$$

where I_{mc} is the minority-carrier current reaching the collector junction from the base. In this treatment β is not simply a constant but an operator which operates upon a varying current γI_e to give I_{mc}. The appropriate form for β will be derived in the following. The current I_{co} is the minority current which would reach the collector junction for zero emitter current. Assume, for simplicity, that this current I_{co} all comes out of the base and neglect space-charge generation and electron current coming from the collector body.

The multiplication factor for holes at the collector junction is M so that

$$I=I_c=MI_{mc}$$

where M is a function of the voltage across the collector junction, V_c. Where it is necessary to use the functional form for M, use the approximation

$$M(|V_c|)=\left[1-\left(\frac{|V_c|}{V_B}\right)^n\right]^{-1}$$

where $|V_c|$ is the absolute value of collector voltage and V_B the absolute value of the breakdown voltage.

If a constant base current I_b flows, the current-voltage relationship between I and V_c may be obtained by noting that

$$I_e=I-I_b$$

so that

$$I=MI_{mc}=M[I_{co}+\alpha(I-I_b)]$$

where α is the conventional alpha for d-c conditions and in general is given by

$$\alpha=\beta_\gamma$$

Dividing I by MI leads to

$$\frac{1}{M}=\frac{I_{co}-\alpha I_b}{I}+\alpha$$

In order to visualize the behavior of this equation, consider cases in which

$$I\geq I_b>I_{co}$$

This corresponds to current in the reverse direction at the collector junction and a flow of majority carriers out of the base, i.e., the base current is trying to reverse bias the emitter junction. Now I is increased from I_b to higher values. If I is equal to I_b, the value of M is

$$\frac{1}{M}=\frac{I_{co}}{I_b}$$

thus, if $I_b=100I_{co}$, for example, M is 100. Now, suppose I is increased to $2I_1$, then M is given by

$$\frac{1}{M}=\frac{I_{co}}{2I_b}+\frac{\alpha}{2}$$

If I_b is equal to $100I_{co}$ and α is near unity, M is reduced to about 2. Thus, increasing I from I_b to $2I_b$ produces a large drop in collector voltage.

Algebraically this change in voltage corresponds to a negative resistance. Thus, in increasing I from I_b to $2I_b$, the change in I_c is $-I_b$ while the change in V_c is from a negative value near $-V_B$ to a much less negative value. Thus a negative change in I_c is produced by a positive change in V_c.

Now, suppose that the switch is closed when the bias currents are such that M is large and I some what greater than I_b. The same bias currents, I_b and I_c, are maintained from the current sources. Let the deviation of the emitter current be δI_e from its previous value. This produces a carrier current across the collector junction given by

$$M\beta\gamma\delta I_e$$

This added current cannot flow out through the bias supplies and acts to charge the two capacitors, namely C and the collector junction capacitance C_c.

Fig. 2 represents the equivalent circuit for the situation under discussion. The transient current δI_e flows around the circuit and produces a current source across the collector junction. (The expressions in parentheses on the figure are discussed in the following section.)

The transient build-up results from the fact that the current source acts to put charge on both C_c and C and that portion which charges C also flows in the emitter lead and enhances the current source. Since a fraction $C/(C+C_c)$ of the current source flows through C, then

$$\delta I_e = \delta I_e \beta\gamma MC/(C+C_c)$$
$$= \delta I_e \beta\gamma M_e$$

where

$$M_e = MC/(C+C_c)$$

is the effective multiplication factor by which unit current of minority carriers entering the collector junction is multiplied in producing current in the emitter lead.

If the bias is such that the device is in the negative-resistance region, then αI_b is greater than I_{co} and

$$\frac{1}{M\alpha} = 1 - \alpha \frac{I_e - I_{co}}{I}$$

so that $M\alpha$ is greater than unity for d-c or low-frequency changes. Evidently, if $M_e\beta\gamma$ is greater than unity, an increase in emitter current δI_e will produce a bigger increase in emitter current and build-up will occur.

The rate of build-up is limited by the delay for diffusion. If the emitter current builds up rapidly, the ratio of collector current to emitter current drops. This effect can be dealt with by considering β to be an appropriate function of the rate of build-up and choosing the rate so that the gain factor $M_e\beta\gamma$ becomes unity:

$$M_e\beta\,(\text{rate of build-up})\,\gamma = 1$$

This equation corresponds to an exponentially growing current. This may be seen as follows: Let δp be the disturbance in $p(x, t)$ in the base layer corresponding to the transient disturbance following the closing of the switch. Let $x=0$ be the collector plane, and $x=-W$ be the emitter plane. Consider

$$\delta p(x, t) = -\exp(\lambda t)\sinh ax$$

This satisfies the continuity equation in the base layer if

$$a^2 = (\lambda + \nu)/D$$

where ν is the reciprocal of the hole lifetime. Furthermore, δp is zero at the collector junction. The ratio of current at collector to current at emitter for δp is β, which is a function of λ, and is

$$\beta(\lambda) = 1/\cosh aW$$

If it is assumed that all the current carried across the emitter is due to injection (and not to space-charge recombination), γ will be independent of the magnitude of the current, but it may depend upon λ since as λ increases, the ratio of diffusion current to carrier density increases at the emitter junction. This dependence of γ upon λ will be neglected here; it may play a significant role in silicon devices, however.

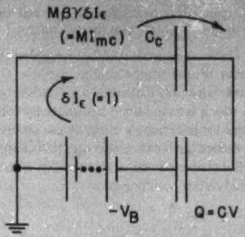


Fig. 2. Equivalent circuit for small and large transients

The equation in δI_e may thus be valid if the build-up constant λ satisfies

$$\beta(\lambda)\gamma M_e = 1$$

For values of γM_e larger than 3, the hyperbolic cosine in β may be represented by its first term and λ becomes

$$\lambda = [D(\ln 2\gamma M_e)^2/W^2] - \nu$$

This is the result quoted previously for the case of $C/C_c = \infty$ (so that M equals M_e) and unit emitter efficiency. The more general expression for $\gamma M_e < 2$ is

$$\lambda = (D/W^2)(\cosh^{-1}\gamma M_e)^2 - \nu$$

The Large-Disturbance Case and the Charge Variables

For the situations with which this paper is chiefly concerned, the current values are so large during the transient phase that the biasing currents are negligible in comparison. Accordingly, it will be assumed in this section that at the same instant that the switch S is closed, the circuit to the collector current source is opened. It will be assumed, initially, that I_b remains unchanged.

The advantage of assuming that the collector current source is reduced to zero is that this permits certain simple conclusions about the relationship between three quantities having the dimension of charge. The first of these quantities is simply the charge Q on the capacitor C of Figs. 1 and 2. First, consider the relationship of Q to Q_{mc} which is defined by

$$dQ_{mc}/dt = I_{mc}$$

Thus, Q_{mc} is, except for an additive constant, the total charge carried by minority carriers arriving from the base at the collector junction. There is a relatively simple relationship between Q_{mc} and Q which is independent of the initial conditions of the discharge and of its progress. To derive this relationship, the relationship between M and Q must be considered.

The battery of potential $-V_B$ has been chosen so that when $Q=0$, V_c is at the breakdown potential and when

$$Q = Q_B \equiv CV_B$$

the voltage across the collector junction is zero. These relationships correspond to neglecting the voltage drop across the emitter which is in any event small, and furthermore, nearly constant since the emitter junction is forward-biased during the transient that is being studied.

It is convenient to express M as a function of Q and to introduce an approximation valid near breakdown in terms of Q_n as follows:

$$\frac{1}{M} = 1 - \left(1 - \frac{Q}{Q_B}\right)^n$$
$$\doteq Q/Q_n \text{ for } Q < Q_n$$

$$Q_n = Q_B/n$$

The first equality involves writing $|V_c|/V_B$ in terms of Q for the expression of the preceding section.

Now consider what occurs when a charge

$$dQ = Idt = \delta I_e dt$$

flows around the circuit of Fig. 2; the equality of δI_e and δI results from disconnecting the current source to the collector. Evidently

$$dQ = \frac{C}{C+C_e} M(Q)dQ_{mc}$$

since the entrance of a charge dQ_{mc} into the space-charge region of the collector results in a charge $M(Q)$ times larger appearing on the two capacitors and a fraction $C/(C+C_c)$ of this appearing on the capacitor C. Thus the relationship

$$dQ_{mc} = dQ/M_e(Q)$$

results where again use the effective multiplication factor

$$M_e(Q) = CM(Q)/(C+C_c)$$

is used. It should be noted that C_c is also a function of Q and this should be thought as being included in $M_e(Q)$.

Thus, Q_{mc} can be expressed as a function of Q by integrating

$$Q_{mc} = \int_0^Q dQ/M_e(Q)$$

where the additive constant in Q_{mc} is chosen so that Q_{mc} and Q both vanish as breakdown is approached.

If the dependence of C_c upon Q is neglected and the analytical approximation for $M(Q)$ is used, then

$$Q_{mc}(Q) = \frac{C+C_c}{C}\left\{Q-\frac{Q_B}{n+1}\left[1-\left(1-\frac{Q}{Q_B}\right)^{n+1}\right]\right\}$$

For small values of Q/Q_B this reduces to

$$Q_{mc} \doteq \frac{(C+C_c)Q^2}{2CQ_n}$$

It is seen that Q_{mc} is less than Q for small values of Q and for $Q << Q_n$ the ratio of Q_{mc} to Q is very small. This means that when Q is small, $M(Q)$ is large, and a small change in Q_{mc} produces a big change in Q.

Since dQ is the charge flowing around the circuit of Fig. 2, it is evident that when Q is small, carriers are being injected into the base layer in amount γdQ to a greater extent that they are leaving at the collector in amount $dQ/M(Q)$. Thus, in effect injected carriers are being stored in the base layer.

If it is assumed that there is a constant emitter efficiency γ, a stored-charge variable Q_s can be introduced by the equation

$$Q_s \equiv \gamma Q - Q_{mc}$$

For the analytic approximation discussed in the foregoing:

$$Q_s = \left(\gamma - \frac{C+C_c}{C}\right)Q + \frac{(C+C_c)Q_B}{C(n+1)}\left[1-\left(1-\frac{Q}{Q_B}\right)^{n+1}\right]$$

For the simplified situation of $\gamma=1$ and $C << C$ this reduces to

$$Q_s = \frac{Q_B}{n+1}[1-u^{n+1}]$$

where

$$u \equiv |V_c|/V_B = 1-(Q/Q_B)$$

The actual charge of minority carriers present in the base layer will not vary directly as Q_s because minority carriers may be lost by recombination within the base layer as well as by emerging as dQ_{mc}. Furthermore, the current I_b of Fig. 1 must be subtracted from the current I flowing into C in order to obtain I_e. During the important part of the transient, the current I_B will be small compared to I and can usually be neglected.

Two important problems can now be considered:

1. What, approximately, is the maximum rate of voltage drop that can be produced by the transient?
2. How does the maximum rate and the amount of voltage drop depend upon C?

These questions can be answered in terms of the stored-charge variable Q_s. First, note that if there is a stored charge Q_s which is maintained substantially constant by injecting as much at the emitter as comes out at the collector, this charge will distribute itself in the base layer in a steady-state configuration with

$$p(x) = -p_0 x/W$$

where P_0 is the density at $x=-W$ the emitter junction. The values of Q_s and I_{mc} corresponding to this are

$$Q_s = qP_0W/2$$

$$I_{mc} = qP_0D/W$$

and the ratio of these two quantities, denoted by ν_s

$$\nu_s = I_{mc}/Q_s = \frac{2D}{W^2}$$

is in effect the number of times per unit time the charge Q_s must be replenished by injection to make up for the loss due to I_{mc} in order to keep Q_s constant.

As will be discussed in a subsequent section, if a charge Q_s is injected and maintained, then the current I_{mc} is somewhat delayed after the injection, in fact by a time approximately equal to $1/6\nu_s$, before it builds up to a maximum of $\nu_s Q_s$.

The maximum rate of voltage drop on the capacitor C will be $M_e I_{mc}/C$. Inspection of the analytic expression for Q_s shows that I_{mc}/C will be increased by increasing γ and by decreasing C_c/C. The optimum case corresponds to the maximum value for Q_s/C which leads to a maximum rate of drop as V_c approaches zero and M approaches unity of

$$dV/dt = \nu_s Q_s/C = \nu_s V/(n+1)$$

As will be discussed in the following, a slightly higher rate of fall may occur early in the discharge waveform while multiplication is still occurring.

Decreasing C compared to C_c leads to increasing Q_{mc} in respect to Q_B and to decreased values of Q_s/C. Thus increasing C increases dV/dt as was pointed out previously.

If γ is small or if C/C_c is not large enough, the voltage across the collector junction may not fall to zero. This conclusion may be reached by considering the expression for dQ_s:

$$dQ_s = \left[\gamma - \frac{C+C_c}{CM(Q)}\right]dQ$$

Unless $\gamma=1$ and $C_c/C=0$, this expression will become negative as Q approaches Q_B and M approaches unity. Negative values mean that there is a net loss of minority carriers stored in the base. If γ is small enough and C_c/C large enough, not only will dQ_s become negative, but Q_s will become zero before Q reaches Q_B. This situation means that minority carriers stored while M was large, get used up until finally none are left in the base and I_{mc} becomes zero.

Thus the extent of the downward excursion may be estimated by finding the value of Q for which $Q_s=0$.

Two refinements shoud be mentioned. The initial transient may start at some voltage corresponding to Q_o and Q_{so}. In this case, the excursion will build up exponentially, as discussed in the previous section, with $M_e(Q_o)$ applying in evaluating λ. It can then proceed only so far that Q_s returns to Q_{so}. Furthermore, if ν, the reciprocal of minority carrier lifetime, is not much less than λ, the excursion will stop sooner.

However, if the collector current generator of Fig. 1 is disconnected, V_c will eventually drop to zero due to the I_{co} term. This represents a very slow drop compared to the rate $\nu_s V_B/(n+1)$ considered in this section.

The next sections will consider the transition from the initial exponential rate of the preceding section to the final rate estimated in this section.

The S(t) Formulation

In this section the discharge-of-the-capacitor problem will be formulated in terms of an integral equation for dQ_s/dt. The final result will involve expressing I_{mc} in terms of dQ_s/dt. Once this has been done the equation for dQ_s/dt is obtained by writing

$$dQ_s/dt = [\gamma M_e(Q_s) - 1]I_{mc}$$

where the first term on the right represents the carrier injection into the base due to I_{mc} entering the space-charge layer and producing current around the circuit and the second term represents loss from the base.

In order to express I_{mc} in terms of Q_s, the function $S(t)$ is introduced. This function is defined in terms of the properties of the base layer and may include drift-transistor effects. $S(t)$ is defined as follows: At $t=0$ inject unit charge across the emitter junction. Thereafter, maintain the boundary condition $p=0$ at the collector. Inject across the emitter a current exactly equal to the collector current I_{mc}. It is evident that $S(t)$ corresponds to a situation in which Q_s increases by unity at $t=0$ and thereafter remains constant.

If Q_s is known as a function of t, then I_{me} can be obtained by considering $Q_s(t)$ to have resulted from a set of step functions at earlier times, denoted by s, such that each step is

$$\dot{Q}_s(s)ds$$

Each such step produces a current at time t which is $S(t-s)$ times this step. Hence

$$I_{me}(t)=\int_{-\infty}^{t} S(t-s)\dot{Q}_s(s)ds$$

This leads to

$$\dot{Q}_s(t)=[\gamma M(Q_s)-1]\int_{-\infty}^{t} S(t-s)\dot{Q}_s(s)ds$$

which will be used as a basis of analyzing the transient behavior for large signals.

If the lifetime is uniform in the base layer with reciprocal lifetime ν, $S(t)$ may be expressed as a product

$$S(t)=\exp(-\nu t)S_o(t)$$

The function $S_o(t)$ corresponds to finding a solution for minority carriers in the base layer satisfying

$$\partial p/\partial t=D\frac{\partial^2 p}{\partial x^2}-\frac{\partial}{\partial x}(\mu pE)$$

where D and μ are minority-carrier diffusion constant and mobility and E is the electric field. The density p must satisfy the boundary conditions

$p=0$ at $x=0$, the collector

and

$$S(t)=q\left(-D\frac{\partial p}{\partial x}+\mu pE\right)_{-W}=q\left(-D\frac{\partial p}{\partial x}+\mu pE\right)_0$$

which states that the collector and emitter currents are equal and equal to $S(t)$. At $t=0$, p must vanish except at $x=-W$ and the integral there must give unit charge for qp.

If $S_o(t)$ and its corresponding $p_o(x, t)$ satisfy these conditions, then it follows that $S(t)$ and $\exp(-\nu t)p_o(x, t)$ satisfy the boundary and initial conditions and also the continuity equation with recombination included.

In the next section, it will be shown how $S_o(t)$ may be calculated for a uniform base layer with no electric field. If a field is present, $S_o(t)$ may have a complicated behavior and may be represented as a series of pulses for short times. For long times $S_o(t)$ may be calculated from a steady state $p_o(x, \infty)$ and its value is then simply

$$S_o(\infty)=\text{current due to } p_o(x, \infty)/\text{charge due to } p_o(x, \infty)$$

For the uniform base layer its value is ν_s, given in the preceding section,

$$\nu_s=2D/W^2$$

The value for a uniform field is given in the next section.

Next, a quantity derived from $S_o(t)$ and its limiting value ν_s will be introduced by the expression

$$\int_0^{\infty}[\nu_s-S_o(t)]dt=f=\nu_s t_s$$

This quantity has no dimensions. It indicates by what amount $S_o(t)$ falls short of reaching its final value ν_s at $t=0$. The integrand vanishes for large t and in effect $S_o(t)$ is negligible for an average time t_s, defined by the foregoing equation, and thereafter $S_o(t)$ equals ν_s.

If Q_s increases at a substantially constant rate for a time large compared to t_s, then $\dot{Q}_s$ may be taken out of the integral for I_{me} by the following steps:

$$\begin{aligned}I_{me}(t)&=\int_{-\infty}^{t}[S_o(t-s)-\nu_s]\dot{Q}_s(s)ds+\nu_s\int_{-\infty}^{t}\dot{Q}_s(s)ds\\&=\dot{Q}_s(t)\int_{-\infty}^{0}[S_o(-r)-\nu_s]dr+\nu_s[Q_s(t)-Q(-\infty)]\\&=\nu_s[Q_s(t)-Q_{s0}]-f\dot{Q}(t)\end{aligned}$$

Thus the current I_{me} is the current due to the added stored charge during the transient except for the charge added during the interval $t-t_s$ to t which would contribute

$$\nu_s t_s\dot{Q}_s(t)=f\dot{Q}_s$$

if $S_o(t)$ rose to its final value ν_s infinitely fast.

This approximation is very inaccurate during the early exponential phase of the first section of this paper. However, as discussed in the second section, Q_s approaches a limiting value and thus $\dot{Q}_s$ must have a maximum. Near the maximum $\dot{Q}_s$ is approximately constant and near that time I_{me} may be estimated fairly accurately from the approximate expression derived in the foregoing.

If the transient starts with a relatively high value of M and nearly negligible value of Q_{s0}, then near the maximum value of $\dot{Q}_s$ one may write

$$\dot{Q}_s=[\gamma M_s(Q_s)-1](\nu_s Q_s-f\dot{Q}_s)$$

This equation may be solved for $\dot{Q}_s$ and from this the following expressions can be obtained

$$\dot{Q}_s=\frac{[\gamma M_s(Q_s)-1]\nu_s Q_s}{1-f+f\gamma M_s(Q_s)}$$

$$\dot{Q}-I=\frac{M_s(Q_s)\nu_s Q_s}{1-f+f\gamma M_s(Q_s)}$$

These expressions will be used later in estimating the characteritistics of the transient behavior.

The Form of S(t)

For a uniform base layer with collector junction at $x=0$ and emitter at $x=-W$, a method of images can be employed to find $p_o(x, t)$ corresponding to $S_o(t)$. If $U(x_o, x, t)=U(x_o)$ denotes a delta function of holes located at $x=x_o$ at $t=0$, which subsequently spreads into the normal gaussian distribution, then

$$p_o(x, t)=2\sum_n nU(-nW)$$

The fact that this gives the same current across all planes of the form $x=nW$ may be verified by adding to $p_o(x, t)$ the series

$$\sum_n U(nW)$$

which shifts $p_o(x, t)$ by W to the right without changing the currents on the $x=nW$ planes.

The gaussian distributions of $p_o(x, t)$ when spread out produce an average density which finally varies as

$$p_o(x, \infty)=-2x/W^2$$

and correctly gives unit numbers of holes between $x=-W$ and $x=0$.

Numerical calculations of $S_o(t)$, which can readily be done by differentiating term by term of the gaussian series, show that the rise of $S_o(t)$ occurs with an average delay of about $1/6\nu_s$ so that the quantity f of the preceding section is about 1/6.

If a strong drift field exists in the base layer, the problem becomes more complicated. The final value ν_s can, however, be evaluated by considering the steady state and is found to be

$$\nu_s(E)=(D/W^2)y^2/(y+e^{-y}-1)$$

where

$$y=qEW/kT$$

is the number of units of thermal voltage across the base layer.

$S_o(t)$ and $S(t)$ are related to somewhat simpler functions $B_o(t)$ and $B(t)$. The B's are the currents which would flow subsequent to unit injection across the emitter followed by zero emitter current thereafter. If S_o^0, S^0, B_o^0, and B^0 denote the operations corresponding to obtaining I_{me} from $\dot{Q}_s$ in the preceding section, then it follows that

Table I. Approach to Maximum Current

M	10	6.70	5	3.30	2.5	2	1.70	1.40
u^3	0.90	0.85	0.80	0.70	0.6	0.50	0.40	0.30
$F(u)$	0.48	0.58	0.64	0.64	0.6	0.54	0.42	0.32
$G(u)$	0.53	0.69	0.79	0.90	1.0	1.04	1.06	1.06

$$S_e^0=B^0_e+(B_e^0)^2+(B_e^0)^3+\dots$$

and

$$S^0=B^0+(B^0)^2+(B^0)^3+\dots$$

In effect each term represents the delayed effect of feeding the current of the previous term back into the emitter.

$B_e(t)$ can be represented by a series of gaussian functions of alternating signs on all planes of the form $x=(2n+1)W$. The corresponding hole density can be expanded in a series of the form $\sin (2n+1)\pi x/2W$. A series of the form $x \cos 2\pi nx/W$ is required at $t=0$ for $S_e(t)$.

Estimate of the Approach to Maximum Current

The result needed from the last section in order to treat the equations of the section on "The $S(t)$ Formulation" was the value of f, which is approximately 1/6.

If it is now assumed the the transient takes place in a time small compared to carrier lifetimes, so that $S_e(t)$ can be used in place of S, and Q_{so} is taken as near zero and $\gamma=1$ and $C_e/C=0$, then the afore-mentioned equations become

$$\dot{Q}_s=[\nu_s Q_B/(n+1)]F(u)$$

$$\dot{Q}=[\nu_s Q_B/(n+1)]G(u)$$

$$F(u)=u^n G(u)$$

$$G(u)=(1-u^{n+1})/[1-(1-f)u^n]$$

$$u\equiv|V_c|/V_B=1-(Q/Q_B)$$

For the typical value of 3 for n the following values are obtained for $F(u)$, $G(u)$, and $M(u)$ where

$$M(u)=[1-u^n]^{-1}$$

For $u=0$, corresponding to zero collector voltage, $M(u)$ and $G(u)$ are unity and $F(u)$ is zero.

Table I shows that $\dot{Q}_s$, or $F(u)$, has a maximum for M about 4. $G(u)$ is seen to exceed unity slightly showing that dV/dt is slightly larger than its final value when M is about 1.4, corresponding to $u=0.66$ or to a drop in voltage of about 1/3 from V_B.

The time required for the current to build up from $G(u)=0.5$ to $G(u)=0.9$ can be estimated as follows: The corresponding values of u are 0.97 and 0.89 corresponding to a change of charge 0.08 Q_B. The average value of $G(u)$ is 0.7. Hence the time is

$$t=0.08Q_B/[\nu_s Q_B\,0.7/(n+1)]$$
$$=0.44/\nu_s=0.22W^2/D$$

It is probable that this estimate is somewhat long for the following reason:

When the current is building up rapidly the correction term obtained by taking $\dot{Q}_s$ out of the integral in the section, "The $S(t)$ Formulation," overestimates the correction. This will lead to slightly larger values for $\dot{Q}_s$ and Q for values of M as large as 10. For $M=10$, the λ constant of the first section is given by

$$1/\lambda=0.11W^2/D$$

corresponding to an increase of a factor of $e=2.73$ in about $(1/9)W^2/D$.

From this it is concluded that the current will increase exponentially to about half its maximum value with an exponential rate of rise of greater than $10D/W^2$ at $M=10$ or about 3% below V_B. In a subsequent time of about $0.2W^2/D$ the voltage will drop to about 11% below V_B and the current to the load will be about 85% of its final value. Peak current will occur at about 33% V_B and substantially constant current will then flow until the collector voltage drops to zero.

Vol.
Pt. I
195

Optimal Utilization and Extension of Interoffice Trunking Facilities

R. E. KALABA
NONMEMBER AIEE

M. L. JUNCOSA
NONMEMBER AIEE

MANY optimization problems in the communications field are characterized by the fact that they involve large numbers of highly interconnected variables. This has forced communications engineers, planners, and managers to resort to various suboptimization schemes, many of which have been of the empirical or trial-and-error types. This situation has in large part been caused by the lack of adequate mathematical and numerical techniques for analyzing and solving such problems.

The rapid development during the past decade of large-scale high-speed digital computing machines and the concurrent growth of the new mathematical tool, linear programming[1,2] has made many multivariable problems, heretofore thought unmanageable, numerically solvable. In an earlier paper[3] the authors gave a general introduction to the potential role of linear programming in the communications field. Here, for the purpose of illustrating its applicability, they show how to handle several interoffice trunking problems of general nature and interest relating to the optimal utilization of existing plant as well as to the minimal cost design of additional facilities. That linear programming formulations are particularly well-suited to these classes of problems is made evident by the fact that additional features such as growth of a system over a period of years, restrictions on equipment availability, etc., are readily incorporated into the models. Several extensions of the basic model will be given to illustrate this point. Thus the inherent flexibility of the models is readily made apparent.

The Model

The system to be considered consists of a set of n points, the stations or central offices, between which pass a set of links. These links are assumed to be bidirectional; e.g., they can be used for communication in either direction. They

---

Paper 58-1245, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 28, 1958; made available for printing September 9, 1958.

R. E. KALABA and M. L. JUNCOSA are with The Rand Corporation, Santa Monica, Calif.

The authors are indebted to R. P. Thompson and S. E. Mikkelsen of the Pacific Telephone and Telegraph Company for bringing this class of problems to their attention and for their interest and encouragement.

could have just as easily been assumed unidirectional links as, for example, in the case of radio links. At each station jumper connections or patches between links are made so that through trunk lines between near or distant stations may be synthesized from links connected in series. It is assumed further that, as a consequence of traffic studies and otherwise, the desired numbers of trunks between all pairs of stations in the network are known.

With this model the problems to be treated here fall into the following two general classes:

1. Given a network of existing links among the n stations and a set of demands for interoffice trunks, how should these links be connected so the maximal possible number of desired interoffice trunks is realized?
2. Given, in addition to the foregoing, a set of unit costs for constructing additional links, how should the existing plant be extended so that increased demands for trunks may be met at minimal cost? This request, of course, is not only for decisions as to what links are to be added, but for decisions as to how all of the links are to be connected together in the extended system.

In many practical problems, especially for larger systems, one may encounter restrictions on the types as well as on the numbers of links that can be patched together in a single trunk without adding any more signal-enhancing equipment than is already in the circuits. For simplicity and to illustrate the methodology, the next section will consider unrestricted systems, i.e., systems for which it is assumed that any set of links joined together in series will provide a satisfactory trunk. The inclusion of restrictions not considered there complicates the problems conceptually and adds greatly to the amount of numerical computation necessary for solution. In a later section the handling of certain types of these restricted systems will be considered.

Unrestricted Systems

Using the afore-mentioned model, first consider a problem of class 1, i.e., of optimal plant utilization. Later problems of class 2, i.e., of minimal cost plant extension, will be considered.

OPTIMAL PLANT UTILIZATION

Let

a_{ij} = known desired number of trunks running between station i and station j to satisfy the needs of station i to communicate with station j

c_{ij} = known number of existing links between stations i and j available for use in trunks running in either direction

x_{ijk} = unknown number of links running directly between i and j which are connected at j to links ultimately to terminate at k

The range of the indexes i, j, and k is the integers 1, 2, . . ., n. Since the links are bidirectional, redundance in the conditions is eliminated by requiring that i be less than j in the definition of c_{ij}. Furthermore, in the definitions x_{ijk} assume $i \neq j$, $i \neq k$ for obvious reasons. Since x_{ijj} is the number of trunks to terminate at j having a link between i and j as their terminal leg, it is clear that one seeks to maximize the total number of completed desired trunks or equivalently the total number of desired terminal links, namely, the quantity

$$\sum_{\substack{i,j \\ i \neq j}} x_{ijj}$$

subject to a number of linear restrictions on the unknown variables x_{ijk}.

First of all, the x_{ijk} are to be nonnegative. Second, the total number of trunks in either direction using the links running between stations i and j cannot exceed the total number of existing links c_{ij} between stations i and j. Last, the total number of trunks emerging from the station i ultimately to be terminated at station j must not exceed the sum of the demand for trunks from station i to station j and the number of trunks entering station i ultimately to be terminated at station j. Consequently, the difference between the total number of trunks passing through station i on their way ultimately to terminate at station j and the total number of trunks originating at station i ultimately to terminate at station j is not to exceed the total number of desired trunks running between stations i and j to satisfy the trunking requirement from station i to station j. It can be shown that this represents a complete accounting of all the restraints on x_{ijk} appearing in the problem.

This leads to the following mathematical problem:

Maximize

$$\sum_{i,j} x_{ijj} \tag{1}$$

subject to the conditions

$$x_{ijk} \geq 0 \tag{2}$$

$$\sum_k x_{jik} + \sum_k x_{ijk} \leq c_{ij}, \quad i<j \tag{3}$$

$$\sum_k x_{ikj} - \sum_k x_{kij} \leq a_{ij}, \quad i \neq j \tag{4}$$

This is a linear programming problem and may be treated by the general techniques available for handling such problems; in particular, the general simplex method of George Dantzig[4] is available for the numerical computation of answers. Problems of the sort considered in this section involving up to a dozen or so stations in the network can be handled by the simplex method using such codes as those developed by W. Orchard-Hays for the IBM (International Business Machines Corporation) *704* machine. It is clear that special computational algorithms for this and related problems will have to be developed if great strides forward in increasing both the speeds of solutions and the size of the networks handled are to be realized. The authors are investigating a specialization of the simplex method for these problems based on the observation that the matrices of coefficients in the system of inequalities 3 and 4 are very sparse, having only $n(n-1) \times (3n-4)$ nonzero elements out of a total of $(3/2)n^2(n-1)^2$ elements, and that these nonzero elements are either $+1$ or -1. Another possibility is to attempt solution from several new points of view; see Bellman,[5,6] Boldyreff,[7] and Ford and Fulkerson.[8]

It frequently happens in practice that linear programming problems do not possess unique solutions, i.e., different sets of values of the variables may yield the same optimal value of the objective function. This is usually the case here. Consequently, it is desirable to add a secondary constraint to the problem that the optimal value of equation 1 is to be attained using a minimal number of the existing links. At present this can be accomplished in two stages, by first solving the original problem stated and then solving a similar problem wherein the restraints, equations 2, 3, and 4, are retained and supplemented by the one additional restraint that the linear form in equation 1 take on the value obtained in the solution of the first stage, and the objective is to minimize

$$\sum_{i,j,k} x_{ijk} \tag{5}$$

The solution of this problem provides that after having satisfied as many of the original demands for trunks as possible a maximal number of spare links will be available for some additional trunks to handle fluctuations in demand, uncertain traffic estimates, etc.

An alternative procedure to obtain an optimal spare solution is to maximize

$$\sum_{i,j} x_{ijj} - \epsilon \sum_{i,j,k} x_{ijk} \tag{6}$$

subject to the condition, equations 2, 3, and 4 of the first problems, and ϵ is a positive fraction. Solutions of this problem will guarantee a number of completed trunks that will exceed the fraction 1-ϵ of the maximum number of trunks that can possibly be computed with a solution of the first problem, and it will do so with a maximum of spares left over. Thus choosing ϵ to be, say, 0.01, the solution of this problem gives at least 99% of the maximum possible number of completable trunks with maximum efficiency. This is not an unreasonable policy to follow since it is doubtful that the dispersion of estimates of the a_{ij} is as low as 1%. Permitting this slight tolerance in achieving the goal of maximum possible satisfaction of the demands for trunks reduces the computation to the solution of one problem instead of two.

A final and very important step in the mechanization of this work is to program the computer to print out exactly what connections are to be made at each station. Thus the process can be made automatic from the determination of the a_{ij} and c_{ij} to the printing out of trunk facility orders to the technicians regarding the connections to be made at each station.

NUMERICAL EXAMPLE

Using the linear programming point of view, numerical experiments have been set up and carried out on several interoffice trunking problems using the Johnniac, one of the Rand Corporation's large-scale high-speed computing machines. A representative example which was considered is the following 7-station optimal utilization problem.

The set of users' demands is given by the matrix

$$(a_{ij})=\begin{bmatrix} 0 & 91 & 19 & 37 & 54 & 62 & 47\\ 38 & 0 & 50 & 5 & 86 & 72 & 39\\ 18 & 12 & 0 & 78 & 43 & 52 & 110\\ 63 & 22 & 9 & 0 & 88 & 16 & 22\\ 25 & 50 & 14 & 69 & 0 & 31 & 75\\ 57 & 84 & 61 & 48 & 53 & 0 & 83\\ 46 & 81 & 71 & 78 & 95 & 30 & 0 \end{bmatrix} \tag{7}$$

The set of available links is given by the matrix

$$(c_{ij})=\begin{bmatrix} 0 & 20 & 200 & 170 & 190 & 15 & 160\\ & & 180 & 180 & 40 & 50 & 65\\ & & & 210 & 175 & 140 & 20\\ & & & & 75 & 180 & 140\\ & & & & & 170 & 30\\ & & & & & & 140\\ 0 & & & & & & 0 \end{bmatrix} \tag{8}$$

The solution is that it is possible to complete 1,921 trunks, a result which may be obtained by hand computation. It is not a simple matter however, to obtain a most efficient or "maximal spare" solution, that is, one for which as many spares as possible remain, especially in such a problem, where the capacities have been chosen so out of keeping with the demands. The result is that the maximal of spares is 11. A set $\{x_{ijk}\}$ providing this solution is

$x_{122}=6$ $x_{225}=49$ $x_{377}=20$ $x_{633}=86$
$x_{123}=45$ $x_{244}=5$ $x_{411}=63$ $x_{643}=84$
$x_{133}=65$ $x_{246}=15$ $x_{415}=13$ $x_{644}=48$
$x_{136}=2$ $x_{256}=23$ $x_{422}=160$ $x_{661}=57$
$x_{142}=40$ $x_{266}=50$ $x_{433}=9$ $x_{656}=55$
$x_{144}=37$ $x_{277}=39$ $x_{455}=75$ $x_{677}=83$
$x_{146}=17$ $x_{311}=56$ $x_{466}=48$ $x_{711}=46$
$x_{155}=81$ $x_{317}=21$ $x_{477}=62$ $x_{713}=46$
$x_{156}=27$ $x_{322}=43$ $x_{511}=82$ $x_{722}=26$
$x_{166}=15$ $x_{342}=14$ $x_{522}=17$ $x_{744}=78$
$x_{177}=68$ $x_{344}=147$ $x_{533}=14$ $x_{766}=25$
$x_{215}=14$ $x_{347}=40$ $x_{534}=69$ $x_{765}=2$
$x_{231}=38$ $x_{355}=92$ $x_{546}=58$ $x_{766}=30$
$x_{233}=50$ $x_{366}=54$ $x_{577}=30$

with all other x_{ijk} equal to zero.

Since this was a pilot attempt, no effort was made to streamline the computations on the Johnniac, but experience gained here indicated that machines such as the IBM *704* could handle problems of this size in a matter of minutes.

OPTIMAL PLANT EXTENSION

Now consider a simple optimal plant extension problem. It will be clear from the following and an earlier paper[3] how generalizations to random and multistage problems are made.

It is desired to augment the existing interoffice trunking facilities in such a manner so as to guarantee that a desired fraction ρ of the required trunks for the coming year will be available. After making estimates of the unit costs for putting in new links between all stations, show how to make the extension at minimum cost.

The notation introduced in a previous section is used with the modification that c_{ij}^* is the number of links between stations i and j in the existing plant and that c_{ij} is the number of links between stations i and j in the new extended plant, which is now an unknown variable to be determined in the solution of the problem. In addition, the generalized annual unit cost of adding links from i to j is denoted by d_{ij}. The problem then takes the form:

Minimize the cost

$$K=\sum_{i,j} d_{ij}(c_{ij}-c_{ij}^*) \tag{9}$$

subject to the constraints

$$x_{ijk}\geq 0,\quad c_{ij}\geq c_{ij}^* \tag{10}$$

$$\sum_k x_{ikj}-\sum_k x_{kij}\leq a_{ij} \tag{11}$$

$$\sum_k x_{jik}+\sum_k x_{ijk}\leq c_{ij},\ i<j \tag{12}$$

$$\sum_{i,j} x_{ijj}\Big/\sum_{i,j} a_{ij}\geq \rho \tag{13}$$

where ρ is the completion factor, i.e., the fraction of desired trunks completed.

This is once again a linear programming problem, and the same general remarks made in the foregoing apply. It should be pointed out here, though, that with the same code parameter studies may easily be carried out. In particular, the dependence of K, the cost of the augmentation, on ρ, the fraction of the desired trunks completed, may be obtained. This is most useful for planning purposes, since there may be some flexibility in the allocation of funds for the contemplated expansion; a curve of K as a function of ρ shows the least possible expenditure to achieve various fractions of desired trunks completed.

CONTINUATION OF NUMERICAL EXAMPLE

Now, reconsider the numerical example which has been given. It was pointed out that for the data given, only 1,921 trunks out of 2,184 desired could be completed using existing facilities. If the unit costs of adding additional links are given by the matrix

$$(d_{ij})=\begin{bmatrix} 0 & 100 & 200 & 100 & 141 & 223 & 200\\ 100 & 0 & 100 & 141 & 100 & 141 & 223\\ 200 & 100 & 0 & 223 & 141 & 100 & 282\\ 100 & 141 & 223 & 0 & 100 & 200 & 100\\ 141 & 100 & 141 & 100 & 0 & 100 & 141\\ 223 & 141 & 100 & 200 & 100 & 0 & 223\\ 200 & 223 & 282 & 100 & 141 & 223 & 0 \end{bmatrix} \tag{14}$$

what is the minimal cost at which links may be added so that all desired trunks may be obtained? In this case the minimal cost turns out to be 36,702.

Several other points were computed, and Fig. 1 is a plot of K versus ρ, the fraction of the desired channels completed. The additional points computed for the curve in Fig. 1 are (0.94, 17,051), (0.96, 23,210), and (0.98, 29,573). As can be

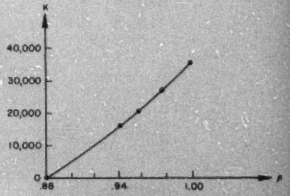


Fig. 1. Minimum cost as a function of completion ratio

seen, the curve is very nearly linear with a slight increase in slope as ρ nears one.

The curve of Fig. 1 gives the ideal minimum cost for which various values of the completion ratio may be obtained, and as such provides a goal which one might try to attain. In practice, due to restrictions not considered here, one would probably have to expend more than K to achieve ρ, so that actually attainable points lie above the curve shown.

EXTENSION WITH RANDOM DEMAND

The problem of designing an interoffice trunk system to accommodate a random future demand is more complicated than the above. A simple approach is to choose the expected values of the demands a_{ij} and use them exactly as one would known demands in the foregoing treatment of these problems. When the demands become known in the future the jumper corrections to be used in the implemented system can be recalculated as the solution of an appropriate optimal utilization problem. This, of course, is based on the assumption that the expected values will not be too far from the demands that actually develop.

The afore-mentioned treatment of the unknown demand situation does not use all the information available, for in general, one may be expected to make reasonable estimates of the actual probability distributions of the demands and not simply their expected values. At the expense of additional computation, one may consider these problems by introducing quantities $p^{(m)}(m=1, 2, \ldots, M)$, the respective probabilities that the matrices of demands $(a_{ij}^{(m)})$ will actually materialize. In addition, one lets the connections $x_{ijk}^{(m)}$ be made if demands $a_{ij}^{(m)}$ materialize. Though other criteria are possible, one might seek the minimal cost extensions to be made so that the expected value of the completion factor ρ will be equal or greater than a smallest acceptable ρ_0.[8] Mathematically the problem becomes that of minimizing

$$\sum_{i,j} d_{ij}(c_{ij}-c_{ij}^*) \tag{15}$$

subject to the conditions

$$x_{ijk}^{(m)} \geq 0,\ c_{ij} \geq c_{ij}^* \tag{16}$$

$$\sum_k x_{ijk}^{(m)} + \sum_k x_{jik}^{(m)} \leq c_{ij} \tag{17}$$

$$\sum_k x_{ikj}^{(m)} - \sum_k x_{kij}^{(m)} \leq a_{ij}^{(m)} \tag{18}$$

$$\sum_m p^{(m)}\Big(\sum_{i,j} x_{ijj}^{(m)} \Big/ \sum_{i,j} a_{ij}^{(m)}\Big) \geq \rho_0 \tag{19}$$

In still further considerations one may desire to plan for the optimal growth of the system over a period of years. Indication of how a multistage process of this sort may be handled are contained in an earlier paper.[8]

Restricted Systems

It is, of course, well-recognized that in many instances links may not be joined together arbitrarily to provide satisfactory service. For telephone plant, for example, such factors as length of the trunks, diameter of the wire, spacing between loading coils, and properties of the repeaters used all enter into a determination of whether or not a trunk composed of various links will be satisfactory. In addition, various equipment incompatibilities may be present. Methods of incorporating these features into the problem will now be considered.

Consider an optimal utilization problem in which various restrictions on the usability of trunks have been laid down. In the general case it now becomes necessary to introduce variables into the problem which are much more complex than the x_{ijk} of the unrestricted problem. In fact, one must now consider all ways of constructing a trunk from i to j and accept or reject them one by one.

Put

$x_{iab\ldots cj}^{\alpha\beta\ldots\gamma}$ = the number of trunks to be used connecting i and j and such that i is joined to a by a link of type α, a is joined to b by a link of type β, $\ldots$, and finally c is joined to j by a link of type γ (20)

If the resulting trunk is satisfactory, admit the corresponding variable into the problem; otherwise simply set it equal to zero. The mathematical problem is to maximize

$$\sum_{\left\{\substack{\text{all satis-}\\ \text{factory}\\ \text{trunks}}\right\}} x_{iab\ldots cj}^{\alpha\beta\ldots\gamma} \tag{21}$$

subject to the conditions

$$\sum_{\left\{\substack{\text{all trunks}\\ \text{joining}\\ i \text{ to } j}\right\}} x_{iab\ldots cj}^{\alpha\beta\ldots\gamma} \leq a_{ij} \tag{22}$$

and

$$\sum_{\left\{\substack{\text{all trunks}\\ \text{including a}\\ \text{link of type}\\ \beta \text{ between}\\ a \text{ and } b}\right\}} x_{iab\ldots cj}^{\alpha\beta\ldots\gamma} \leq c_{ab}^{\beta} \tag{23}$$

where c_{ab}^{β} is the number of available trunks from a to b of type β.

Though this is in general a formidable problem to deal with at present, in cases great simplification is possible; for example, if all trunks consisting of three or more links are considered unsatisfactory, and only one grade of transmission facility is available, one once again has to deal with relatively simple 3-subscript variables. Such problems involving a half-dozen stations and two types of links are within the capabilities of present-day electronic computers using nonspecialized codes.

Other methods are discussed in reference 9.

A number of other considerations, when expressed mathematically, appear in the form of linear equalities or inequalities, or they may be approximated by such expressions. For example, if transmission equipment is in short supply, one may append the condition in the optimal extension problem

$$\sum_{\substack{i,j\\ i\neq j}} (c_{ij}-c_{ij}^*) \leq C \tag{24}$$

where C is the total quantity of such equipment available during the period of plant extension. Various other features which may be incorporated are known and will be reported on in subsequent papers. A very frequent occurrence in the applications of linear programming has been that, as new considerations appear in problems, by the simple addition of more restraints it becomes possible to treat the new problems.

Conclusions

It has been shown that a number of interoffice trunking problems involving both usage and extension can be treated in a straightforward and computationally feasible manner by linear programming methods. To test feasibility two simple representative problems of types A and B were devised for a 7-station unrestricted system as indicated in an earlier section. The rapid, successful solution of these problems on one of Rand Corporation's computers, the Johnniac, as well as the extreme flexibility of the models introduced in this paper, indicates that, as the communications industry turns more and more to large-scale high-speed computers to cope with its large-scale systems problems, linear programming methods will become of increasing importance in carrying out these system-level optimizations.

References

1. AN INTRODUCTION TO LINEAR PROGRAMMING (book), A. Charnes, W. W. Cooper, A. Henderson. John Wiley & Sons, Inc., New York, N. Y., 1953.

2. Linear Programming, R. E. Kalaba. *Pacific Electric Monthly*, Los Angeles, Calif., vol. 1, no. 7, Nov. 1956, pp. 4–8.

3. Optimal Design and Utilization of Communication Networks, R. E. Kalaba, M. L. Juncosa. *Management Science*, Philadelphia, Pa., vol. 3, no. 1, Oct. 1956, pp. 33–44.

4. The Generalized Simplex Method for Minimizing a Linear Form Under Linear Inequality Restraints, G. Dantzig, A. Orden, P. Wolfe. *Pacific Journal of Mathematics*, Berkeley, Calif., vol. 5, 1955, pp. 183–95.

5. On a Dynamic Programming Approach to the Caterer Problem—I, R. Bellman. *Paper P-866*, The Rand Corporation, Santa Monica, Calif., May 1956.

6. Notes on Matrix Theory—XII: Slightly Intertwined Matrices, R. Bellman. *Paper P-917*, The Rand Corporation, Aug. 1956.

7. Determination of the Maximal Steady State Flow of Traffic Through a Railroad Network, A. W. Boldyreff. *Journal*, Operations Research Society of America, Baltimore, Md., vol. 3, 1955, pp. 443–65.

8. Solving the Transportation Problem, L. R. Ford, Jr., D. R. Fulkerson. *Management Science*, vol. 3, no. 1, Oct. 1956, pp. 24–32.

9. On Some Communication Network Problems, R. Kalaba. *Paper P-1325*, The Rand Corporation, Santa Monica, Calif., Apr. 1958.

◆

Discussion

George B. Dantzig (The Rand Corporation, Santa Monica, Calif.): Planners in the communication (and by analogy transport) industries are presented with a challenging new tool for interconnecting and extending units of their system, in an optimal manner. The authors demonstrate that a mathematical model, called the linear programming model, can be formulated that considers simultaneously: (1) the required service between points of the network, (2) the available in-place equipment, and (3) the cost of installing new equipment. Moreover they solved a rather complex example on an electronic computer to demonstrate that models formulated in this manner can be numerically solved.

Experience in other industries, particularly petroleum, indicates that the successful exploitation of this approach requires patience and willingness to begin with models, modest in scope and detail, and eventually to expand the model to reap its full potential.

The communications industry has a rich heritage of planning tools. Hence any new method to compete must overcome the "handicap" of entrenched procedures not present in other industries. However, considering the magnitude of the communication system, the development of this improved planning tool is fully justified.

Comments on the adequacy of the formulation to solve any particular problem of the communication industry are not appropriate since the authors' purpose was only to sketch the possibilities. Actual problems contain a great number of special conditions. Fortunately, linear inequality systems are ideally suited to express commonly encountered conditions including those often approximated by nonlinear functions. Conversely, nonlinear functions (that satisfy certain convexity properties) are easily approximated to any degree of accuracy by linear inequalities.

I would like to comment briefly on recent developments that show promise of solving large-scale systems that might be encountered. Present electronic computer codes are of two types for solving (1) general-purpose problems and (2) transportation or distribution problems. The latter are linear programming problems with a very special structure. The authors' model also has a very special structure that is described as a kind of generalized distribution problem. In this connection Ford and Fulkerson[1] have proposed a method for generating paths through a network for multicommodity distribution problems that have direct application here. Recently the entire approach has been generalized (by Philip Wolfe and the discusser) into what appears to be a fundamental partitioning method which permits the separation of larger systems into component parts.

Reference

1. A Suggested Computation for Maximal Multi-Commodity Network Flows, L. Ford, R. E. Fulkerson. *Management Science*, Philadelphia, Pa., vol. 5, 1958, pp. 97–101.

R. I. Wilkinson (Bell Telephone Laboratories, Inc., New York, N. Y.): The authors have undertaken to apply a new technique, linear programming, to the very old problem of interoffice trunk engineering. Their presentation is thought-provoking and deserves study by those concerned with the economical layout of complicated trunking patterns.

The paper states that in the interoffice trunking of communications systems, engineers have been "forced...to resort to various suboptimization schemes, many of which have been of the empirical or trial-and-error types." And that "This situation has in large part been caused by the lack of adequate mathematical and numerical techniques for analyzing and solving such problems." It is implied that these limitations are about to be lifted by the development of high-speed computers and linear programming "which has made many multivariable problems, heretofore thought unmanageable, numerically solvable."

One is compelled here to register a slight protest, since at least for most telephone interoffice trunk engineering, the "limitations" are not yet very far lifted by this paper. This is due largely to two unrealistic assumptions which the authors have made in order to apply linear programming methods to the problem. And unfortunately there is no apparent easy way to circumvent these assumptions.

First, the authors have assumed that every office will have exclusive use of a number of permanently connected circuits (including 0) to every other office in the area. However, in multioffice trunking plans of the type contemplated, there soon comes a time when some switching, either by built-up end-to-end connections, or by alternate routing through tandems, or both, is more economical than requiring that calls be completed only over permanently assigned direct circuits between each pair of offices.

Second, the paper sets up a radically different criterion of adequacy than has commonly, if not universally, been adopted for the dimensioning of American and foreign telephone systems. The paper proposes a single over-all objective of excellence which may roughly be stated: Form the maximum possible number of desired trunks, and use the minimum number of links to do it. This is an understandably attractive criterion to the engineer of complex systems; it is easy to visualize, it initially appeals to reason, is relatively simple to formulate mathematically, and, perhaps most engaging of all, leads to numerical specification of trunk quantities in a wholly mechanical manner.

However, in local systems most telephone companies adopt as a prime objective the giving of a relatively uniform good grade of service to each customer whatever the destination of his calls. The Kalaba-Juncosa solution says nothing about the grade of service being rendered in the system. Moreover, even if the simple meeting of some selected level of service in the aggregate were accomplished, it would hardly be a sufficient basis on which to allocate equipment and trunking facilities to individual routes. It is easy to show that, even in very elementary system layouts, better choices of trunk plan can be made (in the sense of more load carried and more uniform service given) than are obtained by the over-all optimization program proposed in this paper. There seems little reason to expect that better results would be obtained in more complex networks. Thus, contrary to the authors' introductory criticism, a certain amount of suboptimization appears to be a necessary requirement for attaining the most acceptable service in a given situation.

The relation between load offered to a group of devices and the service given (in the usual sense of proportion of calls lost or delayed, average length of delay, etc.) is seldom, if ever, linear. Therefore, simple optimization by the techniques of linear programming of some linear function of numbers of channels provided cannot generally be counted upon to yield the most satisfactory service—or better, the least unsatisfactory service—to the customers.

In spite of the afore-mentioned shortcomings, it is not inconceivable that the linear programming attack, or some allied method, might yet prove useful in certain communication traffic flow problems comprising a network of origination and termination points. Students of these problems should therefore be familiar with the method.

R. P. Thompson and **S. E. Mikkelsen** (The Pacific Telephone and Telegraph Company, Los Angeles, Calif.): When over 300,000 cable pairs are connected together to form about 150,000 voice paths linking more than 100 buildings, and with approximately 40,000 cable pairs being added each year, the problem of system design and system utilization becomes extremely difficult. This is the size of our problem in interoffice trunking at Pacific Telephone for the Los Angeles metropolitan area. Obviously the combinatorial possibilities are astronomical.

We first met Drs. Kalaba and Juncosa when they gave a talk on a mathematical model of the nation-wide toll switching network. We called their attention to the fact that the trunking in major cities might be a larger and more complex problem than nation-wide toll switching. We asked if they could adapt their mathematical model to this local interoffice trunking problem. This paper is their answer.

Their use of linear programming and the simplex method in their numerical experiments has created interest here in these techniques. In addition to their reference 2, the book, "Linear Programming and Economic Analysis," is useful as an introduction to this subject.[1]

We have been thinking of presenting to a computer a problem in linear programming that might get around the "restrictions" mentioned in the paper. To state the problem will require a little defining of terms.

Interoffice trunking means cable connections between central office buildings. (The word "stations" is used in the paper.) Each pair of conductors has a building at each end. We say connections between two specific buildings lie in a "route." ("Span" is used in some areas and "link" is used in the paper.) When there are many buildings in an area it is not practical to connect each building to all the others directly. We have about 350 routes between approximately 100 buildings. There are several grades, or types, of conductors with an average of about three in each route with a total near 1,000 route-grade combinations.

Traffic studies divide the voice paths needed into groups called "trunk groups." Each trunk group is between a pair of buildings and there may be many between the same pair differing in their requirements as to grade of service or equipment to be used. The required number of trunks is predicted by methods outside of the scope of this particular paper and are assumed to enter this problem as a given list of about 8,000 items. This is the same type of assumption used in the paper.

Since not all buildings are directly connected by routes, a great many of the trunk groups are served by built-up connections of routes in series, by means of jumpers. This is called a design. The design specifies the series connection of route grades, and there may be several designs for each trunk group.

Essentially, the idea is this: Forecasted growth, trunk group by trunk group, is bidding for spare pairs in the route grades called for in the designs for each trunk group respectively. How much of the growth of a given trunk group will work on each of its designs? Designs for different trunk groups will be competing for the same spare pairs in a given route grade. If a design fails in the bid, other alternate designs for the same trunk group will be tried and they may call on a different route or grade where there may be more spares.

This can be answered by linear programming once a criterion is established. This criterion can be the maximum number of trunks that can be connected, as was done in the paper, or the maximal number of spares, which was also done. Another possibility is that use of the plant which will produce the minimum total annual charge. This automatically favors hookups that use the shortest lengths and cheapest grades. Note that we do not solve the restrictions problem, but merely avoid it by predetermined designs each of which is equally suitable for its purpose although the cost may not be the same. This plan is no better than the designs offered into the system. The more designs, the greater possibility of optimizing.

With about 8,000 trunks groups, if only an average of two designs each is offered, this would yield 16,000 total designs. With 16,000 designs and 1,000 route-grades, a formidable linear programming matrix is created. Even with modified simplex methods and the latest superspeed computers, a still simpler method is needed. As Drs. Juncosa and Kalaba say, special computational algorithms will have to be developed. We thank them for their contribution toward solution of interoffice trunking problems.

REFERENCE

1. LINEAR PROGRAMMING AND ECONOMIC ANALYSIS (book), Dorfman, Samuelson, Solow. McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

R. E. Kalaba and **M. L. Juncosa:** We wish to thank all of the discussers for their penetrating comments which add greatly to the value of the paper and which illuminate various problem areas. The objective of subsequent research in this area will be the formulation of more realistic models and at the same time the development of more efficient computational algorithms than exist at present. Furthermore, results in programming under uncertainty, dynamic programming, and quadratic programming should find immediate application.

A specialized procedure for solving certain routing problems, based on the idea of Ford and Fulkerson mentioned by Dr. Dantzig, can be found in reference 1. A still further discussion will appear in a paper of the authors now in preparation.

REFERENCE

1. See reference 9 of the paper.

The Design of a Pile Oscillator for the Ford Nuclear Reactor

J. STONE
NONMEMBER AIEE

W. KERR
MEMBER AIEE

A PILE OSCILLATOR is a device that periodically varies the neutron population in a nuclear reactor. The population change may occur as a localized variation in neutron density, or the density of neutrons throughout the reactor may be caused to vary in a systematic manner. Such a variation may be effected by a periodic oscillation of the multiplication constant, k_{eff}, of the reactor. The required changes in k_{eff} have been achieved in reactor oscillators by mechanical oscillations of either absorbers or reflectors located in or near the core.

The oscillator discussed in this paper produces the required change in k_{eff} by oscillation of an absorber placed near the reactor core. Since it is desirable that the variation in neutron population be sinusoidal in time, the absorber is designed accordingly.

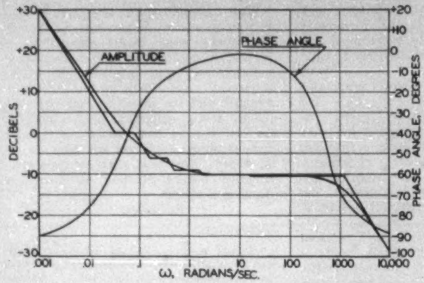
In order to use a pile oscillator, the amplitude of the sinusoidal changes in neutron population must be measured and the phase angle between neutron oscillations and oscillations of k_{eff} must be determined. Appropriate circuitry for making these measurements is also described in this paper.

Pile oscillators have been used for a variety of purposes. Quantities such as neutron absorption and scattering cross sections, delayed neutron characteristics, neutron mean lifetime in the reactor, effective resonance integrals, and the purity of uranium samples have been measured with pile oscillators. Pile oscillators have also been used to determine the dynamic stability of reactor systems and to measure power coefficients of reactivity.

The pile oscillator described in this paper has been designed for the University of Michigan's Ford Nuclear Reactor, a swimming-pool-type research reactor. The oscillator will be used both in connection with teaching laboratories and as a tool for studying the dynamic characteristics of the reactor system.

Paper 58-1329, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 18, 1958; made available for printing September 9, 1958.

J. STONE and W. KERR are with the University of Michigan, Ann Arbor, Mich.



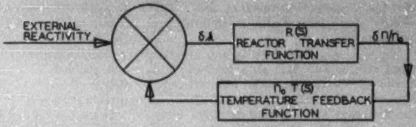


Fig. 1 (left). Calculated transfer function of the Ford Nuclear Reactor

Fig. 2 (above). Internal reactor feedback loop

Nomenclature

n = neutron population, neutrons per centimeter^3

k_{eff} = effective multiplication factor

$\delta k = k_{eff} - 1$

β = fraction of neutrons produced that are delayed; 0.75 per cent of the neutrons present in the reactor are delayed neutrons

τ = mean generation time of prompt neutrons, seconds; (2×10^{-6} second for the Ford Nuclear Reactor)

λ_i = decay constant of the ith group of delayed neutrons, second^{-1}. Five delay groups are considered

C_i = concentration of the ith group of delayed neutrons, precursors per centimeter^3

n_o = steady-state value of neutron population, neutrons per centimeter^3

δn = perturbation of neutron population

C_{i_o} = steady-state value of precursor population

δC_i = perturbation of precursor population

α_i = ith root of the polynomial (see equation 4)

$R(s)$ = reactor transfer function without feedback

$T(s)$ = temperature feedback function

$f(t)$ = signal plus noise

$\psi(t)$ = noise signal

I = value obtained by multiplying the ion chamber output signal by $\cos\frac{2\pi t}{T}$ and integrating for n complete cycles. Experimentally I will be the number of revolutions accumulated on the revolution counter during the n cycles that $f(t)$ is multiplied by $\cos\frac{2\pi t}{T}$. See reference 7

J = value obtained by multiplying the ion chamber output signal by $\sin\frac{2\pi t}{T}$ and integrating for n complete cycles. Experimentally it has a meaning analogous to I

Computation of a Transfer Function for the Ford Nuclear Reactor

A set of equations that can be used to describe the time behavior of the neutron population in a fission reactor operating near critical is given as follows:

$$\frac{dn}{dt}=\frac{(\delta k-\beta)}{\tau}n+\sum_{i}^{5}\lambda_i C_i$$

$$\frac{dC_i}{dt}=\left(\frac{\beta_i}{\tau}\right)n-\lambda_i C_i \qquad (1)$$

The second equation of the set describes the ith delay group precursor population. In a reactor 0.75 per cent of the neutrons originate from certain fission fragments, or precursors, and are not released at the time of fission. These neutrons are called delayed neutrons. Five such groups of delayed neutrons, and thus five precursor groups, are considered in the analysis to be given. The resulting equations form a set of six linear differential equations.

A useful transfer function for purposes of analysis is one that relates changes in neutron population, δn, to corresponding changes in multiplication constant, δk. If δk is assumed to vary with time, then the first differential equation has a time-varying coefficient. In much the same way that one determines the small signal response of electronic amplifiers, the response of the reactor neutron population, δn, to small variations of δk will be found. Let

$$n=n_o+\delta n \qquad (2)$$

$$C_i=C_{i_o}+\delta C_i$$

Here n_o is a steady-state value of neutron population and δn is a small excursion of neutron population about n_o. C_{i_o} and δC_i are similar descriptions of precursor population.

Substituting these expressions for n and C_i into the set of differential equations 1, discarding second-order terms, and using operational notation, produces

$$\frac{\delta n}{n_o\delta k}=\frac{1}{\tau s\left(1+\sum_{i}^{5}\frac{\beta_i}{(s+\lambda_i)\tau}\right)} \qquad (3)$$

Equation 3 can be written in the following form:

$$\frac{\delta n}{n_o\delta k}=\frac{\prod_{i}^{5}(s+\lambda_i)}{s\tau\prod_{i}^{5}(s+\alpha_i)} \qquad (4)$$

where α_i is the ith root of the fifth-order polynomial in the denominator that results from rationalizing the right-hand side of equation 3. See the Appendix.

Using values of β_i and λ_i given in reference 1 and assuming that τ for the Ford Nuclear Reactor is 2×10^{-6} seconds, from equation 4 the transfer function for the Ford Nuclear Reactor is obtained:

$$\frac{\delta n}{n_o\delta k}=\frac{(s+14.3)(s+1.61)(s+0.465)\times(s+0.154)(s+0.0315)}{2\times10^{-6}s(s+364.5)(s+14.21)\times(s+1.425)(s+0.336)(s+0.0775)} \qquad (5)$$

The amplitude and phase angle for equation 5 with $j\omega$ substituted for s are plotted in Fig. 1.

Use of the Oscillator in Determining Reactor Stability

As has been pointed out in the literature, a reactor operating near rated power can be throught of as a feedback system.[2,3] Not only are there feedback mechanisms inherent in the reactor itself, but an automatic control system connected to reactor control rods or to coolant pumps may form additional feedback loops.

As an example of an internal feedback mechanism, consider temperature effects. A small temperature change may vary the moderating and reflecting properties of the reactor core material as well as the geometrical shape. Such changes may introduce a change in k_{eff} which will in turn change n, the neutron population. Fig. 2 is a block diagram of a reactor with internal feedback loops. The closed-loop response of the system of Fig. 2 is

$$\frac{\delta n}{n_o\delta k}=\frac{R(s)}{1+R(s)T(s)n_o} \qquad (6)$$

where, as indicated in Fig. 2, $R(s)$ represents the reactor transfer function without feedback and $T(s)$ is a feedback term.

For experimental investigations it is useful to invert equation 6. There results:

Vol.
Pt. 1
1958

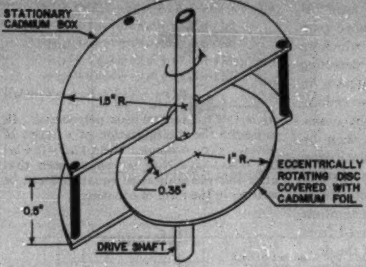


Fig. 3 (left). **Oscillating mechanism**

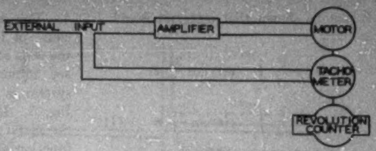


Fig. 4 (right). **Velodyne integrator**

$$\frac{n_0 \delta k(s)}{\delta n(s)} = \frac{1}{R(s)} + T(s) n_0 \tag{7}$$

By methods described in reference 4 the quantities in equation 7 can be determined and the stability of the system investigated. The following procedure may be followed.

1. Operate the reactor at low power so that $T(s)n_0$ will be small compared to $1/R(s)$. Vary δk sinusoidally over a range of frequencies, from 0.001 to 100 cycles per second in the case of the Ford Nuclear Reactor, and measure δn with the circuitry described in the next section of this paper. If the power level is low enough $1/R(s)$ may be computed from these data. The experimental values may be checked with that given by equation 5.

2. Operate the reactor at a higher power level. After allowing the system to reach temperature equilibrium determine the reactor response over the same frequency spectrum as in step 1. As shown by equation 7 these data should allow computation of the quanity $1/R(s)+T(s)n_0$.

3. Instability will exist if for some frequency and some n_0 the quantity $1/R(s)+T(s)n_0$ is equal to zero. This may be investigated by plotting the transfer function $(n_0\delta k)/\delta n$ as a complex number for the frequencies at which data are taken. At a given frequency the vector difference between the transfer function at very low power and at some higher power is the quantity $n_0T(s)$. If this vector can be extended to go through the origin then there exists some power level at which the reactor system will be unstable. This follows from the fact that at a given frequency the phase angle of $n_0T(s)$ is independent of the reactor power level, n_0.

Mechanical and Electrical Design

The Ford Nuclear Reactor uses the BSF (Bulk Shielding Facility) type of fuel elements in a matrix surrounded by graphite reflecting elements. It was felt that it would be most convenient to insert the the oscillating absorber into a position normally occupied by a reflector element.

The design of the oscillator involved computation of the δk introduced by the oscillation of an appropriately shaped absorber in the reactor. Calculations were made using 2-group perturbation theory. These showed that about 0.5 gram of cadmium foil placed in the outer row of graphite reflectors would produce a change in k_{eff} of 10^{-5}; see reference 5. Since derivation of the reactor transfer function assumes a small δn, the oscillator was designed with this in mind. Other calculations showed that the δk produced by the moving absorber of the oscillator would not be much different from that produced by the same quantity of a stationary absorber.[5]

An eccentrically mounted circular disk covered with cadmium foil and rotated with constant angular velocity in and out of a fixed cadmium box was designed to vary k_{eff} sinusoidally. By choosing the eccentricity such that the distance between the center of the disk and drive shaft is 0.35 r, where r is the rotating disk radius, the exposed cadmium area, and thus k_{eff}, will vary harmonically. Fig. 3 shows the mechanical design of the absorber.

For detecting the changes in neutron population, a parallel-circular-plate boron-coated ionization chamber is used.

Both the random nature of the detection process used in the ionization chamber and the randomness of the multiplication process in the reactor introduce noise along with the desired signal. In order to separate the approximately sinusoidal signal from this noise a correlation technique has been used. This is simply the multiplication of the signal by sine and cosine functions of proper frequency. This is accomplished by employing sine and cosine potentiometers driven at the same frequency as the mechanical oscillator. The output signal of the ionization chamber, after amplification, can be represented as

$$f(t) = A\cos\left(\frac{2\pi t}{T} - \theta\right) + \psi(t) \tag{8}$$

where

A = signal amplitude
T = the period of the true signal and of the oscillating δk
θ = the phase shift of the signal with respect to the oscillating δk
$\psi(t)$ = the noise signal

If this function is multiplied by $\sin 2\pi t/T$ or $\cos 2\pi t/T$ and integrated for n complete cycles there results:

$$I = \frac{2}{nT}\int_0^{nT} f(t)\cos\frac{2\pi t}{T}\,dt$$

$$J = \frac{2}{nT}\int_0^{nT} f(t)\sin\frac{2\pi t}{T}\,dt \tag{9}$$

If $f(t)$ is represented as

$$f(t) = A\left(\cos\frac{2\pi t}{T}\cos\theta + \sin\frac{2\pi t}{T}\sin\theta\right) + \psi(t) \tag{10}$$

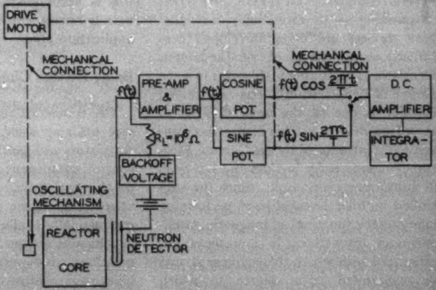


Fig. 5. Detecting circuitry

It can be shown that integrals I and J are now

$$I = A\cos\theta + \frac{2}{nT}\int_0^{nT}\psi(t)\cos\frac{2\pi t}{T}\,dt$$

$$J = A\sin\theta + \frac{2}{nT}\int_0^{nT}\psi(t)\sin\frac{2\pi t}{T}\,dt \qquad (11)$$

The integrals involving the noise terms in equation 11 are on the average zero since the noise is assumed to be nonperiodic. For each oscillation frequency and its corresponding period, T, is found a certain amplitude and a certain angle

$$A = \sqrt{I^2 + J^2}$$

$$\theta = \tan^{-1}(J/I) \qquad (12)$$

The operations described by equations 9 are achieved by the use of the sine and cosine potentiometers.

The integrator chosen is a type that has been used in pile oscillator instrumentation by Breton.[6] It consists of a d-c split-field motor whose angular velocity is proportional to the applied voltage. The angle of rotation of the motor shaft over a certain time interval is thus proportional to the integral of the signal imposed on the field control winding. On the same shaft are connected a revolution counter and a tachometer. The output voltage of the tachometer is fed back to the motor as shown in Fig. 4. This type of integrator is used because of its excellent performance over the wide range of frequencies in which we are interested.[7] The entire detecting system is shown in block diagram in Fig. 5. In the experimental procedure, the signal $f(t)$ is multiplied by the cosine function for 10 minutes and then by the sine function for a similar time. By noting the number of revolutions recorded on the revolution counter in the integrator, the phase and magnitude of the δn signal can be ascertained from equations 12. The frequency of the driving function δk can be changed by changing gear ratios (not shown in Fig. 5) and driving motor speed. As can be seen from Fig. 5 the sine and cosine potentiometers as well as the mechanical oscillator all move at the same frequency.

Appendix

Equation 3 may be written

$$\frac{\delta n}{n_0\delta k} = \frac{1}{s\left[\tau + \frac{\beta_1}{s+\lambda_1} + \frac{\beta_2}{s+\lambda_2} + \frac{\beta_3}{s+\lambda_3} + \frac{\beta_4}{s+\lambda_4} + \frac{\beta_5}{s+\lambda_5}\right]}$$

Rationalizing the denominator of this equation results in the following fifth-order polynomial:

$$\begin{aligned}\tau[(s+\lambda_1)(s+\lambda_2)(s+\lambda_3)(s+\lambda_4)(s+\lambda_5)+\\ \beta_1(s+\lambda_2)(s+\lambda_3)(s+\lambda_4)(s+\lambda_5)+\\ \beta_2(s+\lambda_1)(s+\lambda_3)(s+\lambda_4)(s+\lambda_5)+\\ \beta_3(s+\lambda_1)(s+\lambda_2)(s+\lambda_4)(s+\lambda_5)+\\ \beta_4(s+\lambda_1)(s+\lambda_2)(s+\lambda_3)(s+\lambda_5)+\\ \beta_5(s+\lambda_1)(s+\lambda_2)(s+\lambda_3)(s+\lambda_4)]\end{aligned}$$

α_i is the ith root of this polynomial. Π_i indicates the multiplication of a series of terms. For example, $\Pi_i{}^3(s+\lambda_i) = (s+\lambda_1)(s+\lambda_2)(s+\lambda_3)$. Now it can be seen that after this rationalization equation 3 can be written in the form of equation 4.

References

1. DELAYED NEUTRONS FROM FISSION, D. J. Hughes, J. Dobbs, A. Cahn, D. Hall. *Physical Review*, New York, N. Y., vol. 73, Jan. 15, 1948' pp. 111–24.

2. CONTROL OF NUCLEAR REACTORS AND POWER PLANTS (book), M. A. Schultz. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

3. *Transactions*, Professional Group on Nuclear Science, Institute of Radio Engineers, New York, N. Y., Mar. 1957.

4. THE EFFECT OF TEMPERATURE COEFFICIENTS ON REACTOR STABILITY AND REACTOR TRANSFER FUNCTION, R. Siegel, H. Hurwitz, Jr. *KAPL 1138*, General Electric Company, Schenectady, N. Y., Mar. 1955.

5. A PILE OSCILLATOR PROPOSAL FOR THE FORD NUCLEAR REACTOR, F. Bremsak, B. Cimbleris, J. Stone. University of Michigan, Ann Arbor, Mich., 1957.

6. TESTING OF MATERIALS BY THE OSCILLATION METHOD AT THE REACTOR AT CHATILLON, D. Breton. "Progress in Nuclear Energy Series", Reactors, McGraw-Hill Book Company, Inc., New York, N. Y., 1956, pp. 217–33.

7. THE VELODYNE, F. C. Williams, A. M. Utley. *Journal*, Institution of Electrical Engineers, London, England, 1946.

The Controlled Rectifier: Key to the Continuing Control Renaissance

J. D. HARNDEN, JR.
ASSOCIATE MEMBER AIEE

THE SEMICONDUCTOR art is expanding at an exceedingly fast rate. New devices are being developed continually so that the use of even the famous transistor is beginning to decrease in some areas of application. The controlled rectifier is one of the most exciting of these new devices. This device was commercially introduced by the General Electric Company early in 1958 and has been in production ever since. Since the controlled rectifier is capable of performing many of the functions of magnetic amplifiers and thyratrons, it represents another tool with which the engineer should be acquainted. The degree to which this is accomplished will determine the new horizons which are possible in the application area.

Basic Device Characteristics

The controlled rectifier is a silicon device utilizing some of the properties of transistors and rectifiers to produce a new control component somewhat analogous to the thyratron. The device circumvents some of the problems surrounding commercially available power transistors, in achieving greater voltage ratings, faster switching times, and in being capable of higher temperature operation. The basic device consists of four regions, p-n-p-n, as shown in Fig. 1. While the device appears to be somewhat like a hook transistor in that both have four regions, its method of operation and application is markedly different, and it must be treated as an entirely new device. The basic anode-to-cathode characteristic is also shown in Fig. 1, which reveals the fundamental switching action.

In addition to the diode method of control, by far the greatest potential surrounds the use of the third element for triggering. Using this gate method of control, the controlled rectifier has a minimum power gain (ratio of output power to input power) in excess of 150,000 with switching times of the order of 1 microsecond. This power gain assumes

---

Paper 58-1234, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing August 26, 1958.

J. D. HARNDEN, JR., is with the General Electric Company, Schenectady, N. Y.

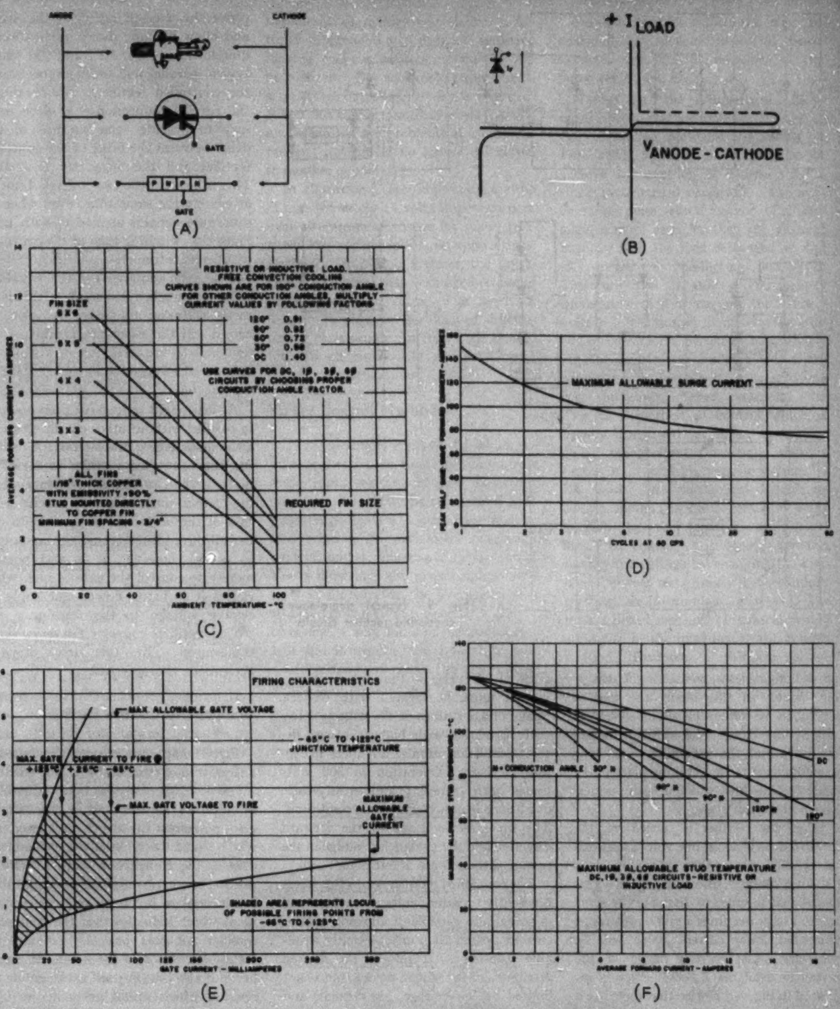


Fig. 1. ZJ-39A controlled rectifier characteristics. A—Gate. B—Anode-cathode. C—Required fin size. D—Maximum allowable surge current. E—Firing characteristics. F—Maximum allowable stud temperature

continuous gate drive. For pulse methods of triggering, the power gain should be re-evaluated, depending on the pulse width. The device exhibits a "deionization time" which is somewhat analogous to that which takes place in thyratrons. This time is in the order of three times the turn on time. The possibility exists that the controlled rectifier can be turned off, but it may not be desirable to design the unit to allow utilization of this characteristic. However, with this potentiality, the device would become very analogous to the ideal switch in that a pulse could turn the rectifier on, allowing it to remain in a highly conductive state in the absence of control power, and to be turned off, reverting to its open condition by applying negative signal information for a short period of time.

It must be emphasized that the device does not require continuous signal excitation as does a transistor. The transistor is a proportional device which has im-

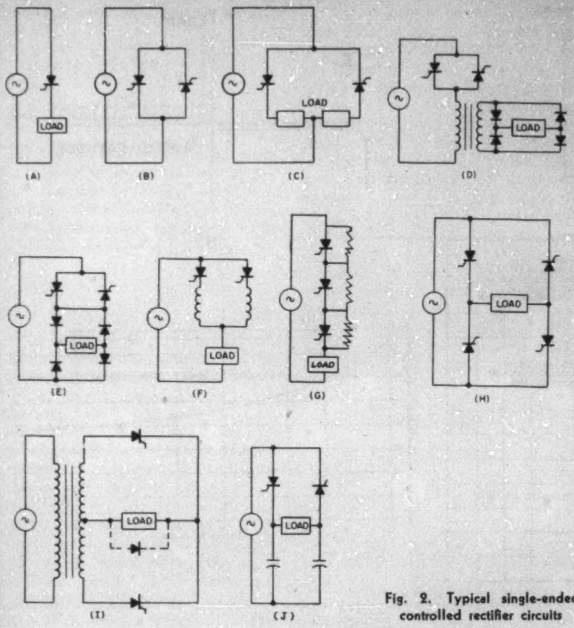


Fig. 2. Typical single-ended controlled rectifier circuits

proved properties when used in the switching mode. In this mode, the unit is forced to be either closed or open by driving the base. The switching losses are large and the switching times long in comparison to those taking place in the controlled rectifier, which actually undergoes a change of state.

When the rectifier is turned on, the forward drop is about one volt or approximately 20% greater than the voltage drop for an ordinary silicon rectifier. This means that in order to dissipate the same amount of power from a given package, a somewhat lower current value will be necessary. The current rating of the presently available *ZJ-39A* series is indicated in Fig. 1. Notice that there are a number of current ratings, depending on the application. The allowable current depends on the size of the unit, i.e., the junction area, the method and effectiveness of cooling, and the per cent ON time of the current wave form.

Like any power semiconductor device, the time constant of the junction is of importance during transient conditions, while the thermal heat sink must take care of steady-state losses. The time constant of the junction is less than a half-cycle of 60 cycles so that transient protection must be treated with respect. Peak voltage ratings of 400 volts are presently available with higher voltage designs paralleling similar advances in regular rectifiers. Operation to 500 volts under both transient and steady-state conditions is possible for some conditions. The breakover voltage in the forward direction and the avalanche voltage in the reverse direction are about equal.

The controlled rectifier seems to possess inherently greater voltage ratings and is potentially capable of allowing greater current ratings than any presently available transistor. It retains the desired attributes of the silicon device, but is not limited by low-voltage, low-current, and high-saturation resistance as is common for commercially available transistors. It is obviously only a matter of time before controlled rectifiers will be available in ratings similar to those of ordinary silicon rectifiers, thereby competing with and replacing saturable reactors, magnetic amplifiers, thyratrons, and ignitrons in many applications.

It is possible to parallel and also to connect units in series, as well as using them in multiphase connections to improve the over-all rating of the system and allow meeting many application requirements. Figs. 2(F) and (G) show a typical parallel and series connection for the controlled rectifier. Notice that in the parallel connection it is often necessary to provide some method of time delay between the firing of one controlled rectifier and the firing of other units. This is because the two units will not fire at exactly the same time, even when the same gate signal is applied to both units. Since the switching time of the controlled rectifier is extremely fast, the number of volt-seconds required for the equalizing inductors is relatively small, if this is the method chosen to delay the switching action. If this reactance is omitted, the first unit to fire could act as a short circuit on the second unit, preventing its turning on.

At least eight units have been operated in parallel without other circuit elements, after some degree of selection. Connecting the controlled rectifier in series requires a small amount of resistance-shunting of each unit, or a very careful selection of the units. It is possible to control all three units to cause the breakdown to occur; however, it is also possible merely to control the gate of one, which causes it to break down, and apply excessive voltage to the remaining controlled rectifiers connected in series across the source. They then break down on their diode characteristic.

Competitive Control Device

Fig. 3 compares a saturable reactor and a thyratron of comparable rating with the *ZJ-39A*. Not only is the controlled rectifier far smaller and lighter, but it also possesses the other characteristics which have been briefly outlined. Of course, the comparison with a self-saturating magnetic amplifier of equal rating would need to involve a rectifier complement which indicates that the controlled rectifier not only basically replaces the saturable reactor in the magnetic amplifier, but also that in itself it can act as the rectifier which would normally be used in such an application. The controlled rectifier is capable of a high degree of efficiency since only one volt is lost or not delivered to the load for supply voltages approaching 400. In a magnetic amplifier, there are several factors which prevent the full application of voltage to the load consisting of the resistance of the gate windings and the reactance of the inductive circuit. The magnetic amplifier must have sufficient turns to prevent saturation of the core on the applied volt-

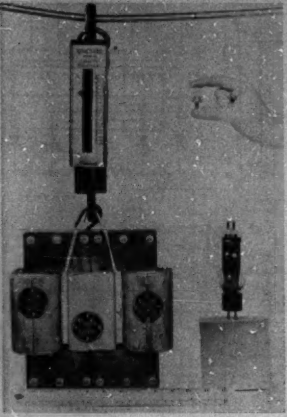


Fig. 3. Amplistat reactor, thyratron, and ZJ-39A controlled rectifier of approximately equal power rating

age, and these turns must be capable of carrying full load current without overheating. These two factors cause the amplifier to be a relatively large device. The gain of the magnetic amplifier is, of course, variable depending on the amplifier's response time, frequency, etc. The controlled rectifier, on the other hand, is not capable of interchanging gain for response, being basically a fixed-response device. For all intents and purposes in industrial control systems, the transfer function of the controlled rectifier may be considered merely as a gain stage with no transport lag, phase shift, or ramp function. As the carrier frequency approaches the inherent switching time of the device, there is some transport lag which should be considered.

All magnetic methods of amplification need not be static. An amplidyne is an example of a rotating amplifier. Two virtues of the amplidyne are that the output varies in magnitude, (i.e., it is not phase-controlled) and that the output is basically push-pull or polarity-reversing. In a sense, an electromagnetic relay is also an amplifier. In addition to size, weight, and environmental specifications, general relay requirements are for lowest input power, highest contact rating, and fastest operating speeds. Clearly, a switch of the electromagnetic type is incapable of matching the response of the inertialess controlled rectifier. The minimum gating power is lower than many so-called sensitive relay requirements at present, and as the family of ratings materialize, sensitivity approaching the most sensitive moving coil contact-making meters will be realized. For each pole requirement, a separate controlled rectifier is required. When this device is used as a relay, shock and vibration are problems of a different nature, and life is no longer calculated in number of operations.

In thyratrons, the high-voltage drop of approximately 15 volts imposes a poor over-all system efficiency for many relatively low-voltage industrial applications. Certainly for 28-volt systems, the thyratron is at a great disadvantage because of its high forward drop. The need for filament supply power, delayed application of plate supply, and relatively limited life are all necessarily associated with thyratron applications.

Power Control Methods

There are many methods by which the voltage across a load may be varied from some input voltage source. Generally speaking, these may be categorized as 1. phase-angle control, 2. time base modulation control, 3. mechanical magnitude control, and 4. impedance ratio control. Fig. 4 illustrates the phase-angle control method for varying the average voltage to a load from a given a-c carrier. This is, of course, a very familiar method of control and it results in a typical transfer characteristic of input versus output which may be expressed by

$$I=\frac{E}{R}\,\frac{1+\cos\alpha}{2}$$

Further, such a waveform may be analyzed for various ratios of peak, average, and rms, as functions of the firing angle; see Fig. 5.

Notice that the symmetrical phase-angle control method requires the use of two controlled rectifiers, as shown in Fig. 2(B), whereas it is possible to achieve this method of control with a single symmetrical transistor. The synchronous pulse width method of phase-angle control requires that the element in series with the load be turned on and off during a conducting half-cycle. Therefore this is not a particularly common method of varying the average voltage. The time base modulated methods of control are becoming increasingly popular due to the availability of sharp saturating devices such as square-loop core materials, Zener rectifiers, etc. This is a method of control which is usually construed to mean generation of an independent time base, thereby producing a variable average voltage to the load. The mechanical magnitude control method is quite common, being typified by the variable transformer idea. Of course, it is equally possible to vary the voltage by changing the coupling between the primary and secondary windings in a number of ways. The least utilized method of control would appear to be the impedance ratio control method. Examples of this are voltage doubler circuits and circuits using a resonant type of variable voltage source. While single-phase representations have been considered, it is obvious that the

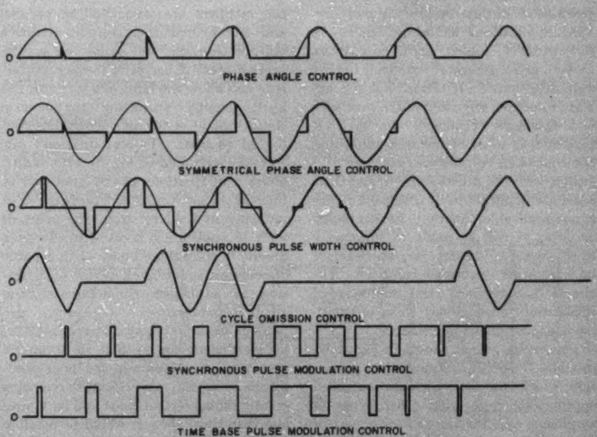
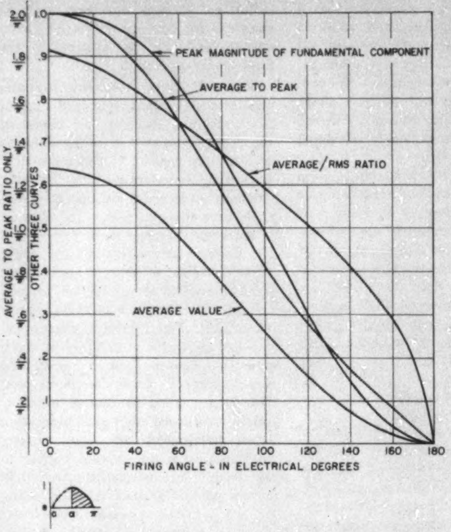

0
PHASE ANGLE CONTROL
0
SYMMETRICAL PHASE ANGLE CONTROL
0
SYNCHRONOUS PULSE WIDTH CONTROL
0
CYCLE OMISSION CONTROL
0
SYNCHRONOUS PULSE MODULATION CONTROL
0
TIME BASE PULSE MODULATION CONTROL

Fig. 4. Methods of power control



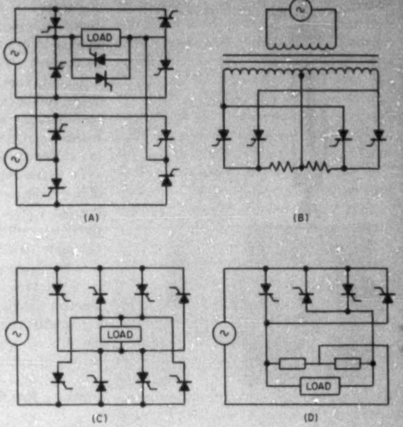


Fig. 5 (left). Theoretical controlled rectifier waveform analysis

Fig. 6 (above). Typical push-pull controlled rectifier circuits

multiphase techniques are equally applicable with the use of the same principles.

General Circuitry

To attempt to list all the possible applications of the controlled rectifier would be almost impossible. However, it is possible to outline briefly the more obvious applications, including such things as control of heaters, control of lamps, motor and generator control, static switching, static excitation, pulsers, circuit protection, servo amplifiers, inverters, and frequency changers. In any application, it must be borne in mind that the controlled rectifier can act only as a switch, i.e., it is either on or off, and there are a few applications in which the device is not applicable directly. In this paper, an attempt will be made to summarize briefly some of the circuitry which might be used with controlled rectifiers. The circuit potentialities and methods of application follow closely those of thyratrons and magnetic amplifiers.

The bridge circuit shown in Fig. 2(H) provides inherent commutation protection, which is as desirable with controlled rectifiers as it is with other types of amplifiers. In the center tap circuit of Fig. 2(I), an auxiliary diode is usually required. In the case of normal rectifier circuitry, the peak inverse voltage ratings required in the circuit connections will follow and parallel directly the requirements in normal rectifier circuitry. Fig. 2(B) is of additional interest because it provides some degree of inherent protection for the semiconductors in the presence of abnormal voltage pulses and surges on the voltage distribution system. Since the rectifiers are connected in parallel and in an inverse direction, any excessive voltage transient will cause one or the other to break down on its diode characteristic and allow the fault to pass on to the load, thereby protecting the reverse characteristic of the nonconducting controlled rectifier. Typical push-pull circuits are indicated in Fig. 6, while multiphase configurations are diagrammed in Fig. 7. The controlled rectifier is a radically new device from the point of view of application because it affords a duality of control methods. It is capable of being gate-controlled, as well as possessing an avalanche characteristic from anode to cathode.

As an example, it would be possible to utilize this device so that in a control system the breakdown of the diode characteristic would take place for voltages greater than a certain value and for values less than this; then, it would be possible to gate control the output. Thus, this device could be used as a limiting element in a feedback system for clipping and control purposes. A simple relaxation oscillator using the diode characteristic is shown in Fig. 8(A). Normally, in such a circuit, the gate would not be used. However, data indicate that by using the gate and allowing negative current to flow, the breakover can be changed, thereby changing the frequency of oscillation.

The circuits shown in Fig. 8(B) and (C) are very similar in nature. These two circuits operate from a d-c source and provide alternating conduction to the split load which can either be resistive or consist of a transformer. In operation, basically, the controlled rectifiers are alternating in their conduction half-cycle, depending on the nature of the input information. In (B), successive pulses cause the load current to flow in alternate branches of the split load. Once triggered into a conducting condition, no further gate power is required and no further change in load will take place until another pulse is applied to the system. In Fig. 8(C) this characteristic is utilized in a simple parallel-inverter circuit. The power requirement for the synchronizing signal is small in this type of circuit. While it is a simple inverter circuit, it is relatively restricted as to its usefulness.

Of course, more complex circuits will

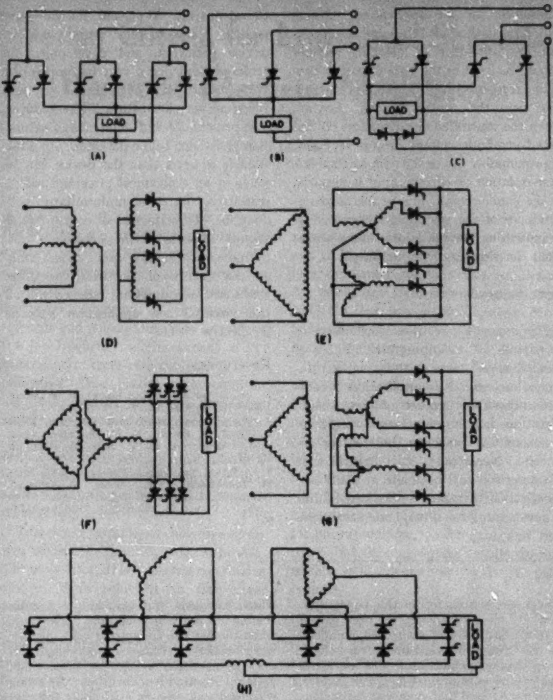


Fig. 7. Typical multiphase controlled rectifier circuits

find wide utilization of controlled rectifiers for improving inverter performance. Notice that this method of inversion depends on external circuit parameters for commutation in contrast to the inherently commutated transistor-saturating core type of inverter. Since the controlled rectifier is a high-power high-voltage device, applications to inverters of high power rating in all phase combinations is an ideal application. Already a number of high-power inverters have been constructed. Indications are that higher frequency operations will be possible than previously with either transistors or thyratrons because of the inherently faster deionization time. In the broad sense, inversion is but a small segment of a technological area, encompassing frequency changers and converters as well. Likewise, many systems cannot be uniquely classified since they may combine several modes of operation, depending on load conditions. In some instances, it may be advantageous to combine a converter with an inverter to achieve best over-all performance with savings in inverter size and weight.

Gating Circuitry

The circuitry necessary for triggering controlled rectifiers is a function of the application requirement and the performance desired. It is possible to use semiconductor preamplifiers in any number of circuits for triggering, phase shift control, etc., in addition to normal-phase shifters, firing circuits, and direct currents The gating action can also be produced by direct injection of energy (such as light) into the crystal itself. In many applications, it appears that the use of magnetic circuitry in the gate circuit will be advantageous because of the isolation which the magnetic circuitry provides inherently, and the large range of phase-angle control and large ratio of maximum to minimum currents obtainable with the use of square-loop core materials. In addition, magnetic devices provide an ideal mechanism for amplifying d-c signals down to relatively low levels such as 10^{-8} watts, or lower, over wide ambient and environmental ranges. The actual wavefront of the load current will usually be faster than that of the gating device, due to the fast firing time of the controlled rectifier. It should also be emphasized that the controlled rectifier imposes no further delay on the over-all system performance. Virtually all of the delay, phase shift, and dead time will be contributed by whatever device is used to gate the controlled rectifier.

In contrast to most conventional magnetic amplifier circuits which produce full output in the absence of signal, the controlled rectifier is a fail-safe device; i.e., in the absence of a signal, the output is zero. The on-to-off ratio of the controlled rectifier is exceedingly high. The average minimum holding current is approximately 10 milliamperes with peak load currents reaching 30 amperes or more. The minimum holding current is analogous to the conduction voltage drop in the thyratron since, if the current falls below this value, the controlled rectifier will revert to the open condition. In the open condition, the leakage currents in both the forward and reverse directions will be approximately the same as those found in a silicon power rectifier of comparable rating. Notice also that this as

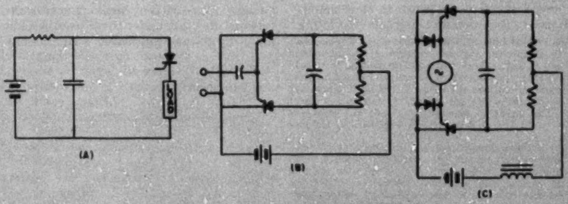


Fig. 8. Typical d-c controlled rectifier circuits

well as other parameters will vary with temperature as does a silicon rectifier. Unlike the magnetic amplifier, exciting current is not required to flow prior to the actual beginning of conduction. Full control is retained until the actual time of triggering. However, there is a finite leakage current which flows in both directions which might be considered somewhat analogous. The amount of leakage depends on the quality of the unit, the temperature, etc.

The gate signal must be greater than approximately one volt to overcome the inherent voltage drop of the lower p-n region of the controlled rectifier. The other requirement is that sufficient current be injected into the lower positive region to cause the avalanche to produce itself at the inner n-p region in the 4-layer construction. It requires approximately 15 milliamperes for this triggering action to take place. Certain precautions should be taken with regard to the gate circuit in order to prevent application difficulties. First, the lower p-n region is a relatively poor rectifier with a low forward-to-reverse ratio of resistance (about 4 to 1). Therefore, for a-c applications, during which the cathode will become positive, a limiting resistance or blocking rectifier should be used in the gate circuit to prevent excessive current flow out of the gate lead into the control circuit. Second, when the controlled rectifier turns on, the gate-to-cathode voltage drop is very low, approximately 0.2 or 0.3 volt, so that further isolation or current limit is required in the gate circuit. The temperature dependence of the gate is indicated in Fig. 1, indicating the need for a sharp wavefront for widely varying ambients.

Conclusions

The controlled rectifier would seem to be capable of revamping any number of control applications with a marked degree of success. New applications heretofore impossible because of size, shock, vibration, heater power, and weight restrictions will challenge the creative engineers. Necessarily, it is natural that a comparison will be made in many applications with regard to the capabilities of power transistors, magnetic amplifiers, thyratrons, etc. At present, the controlled rectifier seems to offer more range for device perfection and development in terms of achieving a very high operating voltage and large current ratings, and therefore very low saturation resistances. The present *ZJ-39A* device has a saturation resistance below 0.2 ohm. In addition, it appears that the device can be made at an economical price and with a reasonable degree of uniformity in production. These factors of course, are of utmost importance in choosing a new component for circuit application. Only as the family of controlled rectifiers grows and new units are released will all the potential new application horizons be clearly visible.

References

1. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

2. Solid State Thyratron Switches Kilowatts, R. P. Frenzel, F. W. Gutzwiller. *Electronics*, New York, N. Y., Mar. 28, 1958, pp 52–55.

3. P-N-P-N Transistor Switches, J. L. Moll, M. Tanenbaum, J. M. Goldey, N. Holonyak. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, no. 9, Sept. 1956, pp. 1174–82.

◆

Discussion

F. W. Gutzwiller (General Electric Company, Clyde, N.Y.): Mr. Harnden admirably achieves the twofold objective of introducing a new type of semiconductor power device to the AIEE and discussing in detail some of the widespread technical ramifications and effects that this new device will have on the uses and control of electrical power. He emphasizes the outstanding advantages of the silicon controlled rectifier over other types of devices that control relatively large blocks of power by switching action. In comparing the controlled rectifier with other devices, it is worth recognizing that the downward price trend of semiconductors counteracts general economic forces that tend to raise the cost of other types of devices. This economic factor combined with the advantages listed in the paper can be reasonably expected to make the controlled rectifier the key that unlocks the tremendous potential for more refined, precise control of electric power in the consumer and commercial fields in some similar degree to that already being achieved in the industrial and military fields.

Since the submission of this paper, substantially more experience has been gained in series operation of controlled rectifiers. The variation of forward breakover voltage with junction temperature can be quite large in devices of present-day manufacture. It has been found that even with meticulous matching of forward blocking characteristics and the other parameters that affect junction temperature, satisfactory series operation without external forcing means requires such drastic voltage derating to prevent misfiring as to make this approach impractical. On the other hand, the use of shunt resistors for forced voltage sharing, as suggested by Mr. Harnden, has proved very satisfactory.

In typical rectifier applications, over-all circuit efficiency is reduced by less than 0.3% by the addition of shunt resistors.

Because of the variation in forward breakover voltage just discussed, it has also proved generally unsatisfactory to fire a series string of controlled rectifiers simply by controlling the gate of one of the cells. Such a string must be able to withstand full circuit voltage without breaking over in the forward direction under conditions of maximum junction temperature which occurs when maximum load current flows under the highest ambient conditions. However, when the devices are started cold, the forward breakover voltages of individual cells may be as much as two or three times greater. Under these conditions the string will not break over by controlling the gate of only one cell. For reliable control, multiple secondaries on a small pulse transformer can be used to feed an electrically isolated control signal to each of the individual cells.

In measuring the forward conductivity of the controlled rectifier, Mr. Harnden uses the transistor term "saturation resistance." Although such a term may have some significance over a limited current range, it is worth noting that the controlled rectifier does not have a characteristic approaching true resistance until load current levels reach several times normal. The *ZJ39A* device described approaches a dynamic resistance of approximately 0.01 ohm at 100 amperes. This low level of dynamic resistance at high current levels is a unique advantage over power transistors of the same general current rating, and it has particular significance in pulse-type applications.

Design Criteria for Low-Level Second-Harmonic Magnetic Modulators

E. J. KLETSKY
NONMEMBER AIEE

MAJOR difficulties are encountered when d-c amplifiers are used to amplify low-level low-frequency signals. These difficulties are mainly problems of zero-drift and noise. One of the most successful techniques in circumventing the drift and noise problems is to convert the low-frequency intelligence to a frequency permitting conventional a-c amplification. The means for accomplishing the frequency conversion have been mainly mechanical: choppers, vibrating capacitors, and similar devices. These devices, by their very nature, must be operated at relatively low sampling rates and are subject to the usual failures of moving equipment.

This paper presents an analysis and design criteria for the magnetic modulator. The modulator, as a frequency-conversion device, offers solutions to the aforementioned problems. It converts low-level low-frequency signals into amplitude-modulated suppressed-carrier outputs. The device has extremely low inherent zero-drift, wide frequency range extending down to direct current, and considerable power gain. Along with these electrical characteristics, the modulator has no vacuum tubes, no moving parts, nor does it require critical adjustment; hence it is extremely rugged and practically maintenance-free.

In spite of the modulator's desirable features, it has not found widespread application. A glance into the history of the magnetic modulator indicates why its striking characteristics have not been previously exploited, and what is needed in order to exploit these characteristics more fully.

The basic operating principle of the magnetic modulator has been known for many years. In the early 20's the Bell Telephone Laboratories carried out preliminary investigation of the modulator.[1] However, the lack of magnetic materials with high permeability, and the advent of a new and interesting device, the vacuum tube, caused research on the magnetic modulator to bog down. For the next 15 or 20 years, the device lay dormant.

During World War II the modulator was re-examined. With the aid of new magnetic materials, it was developed into an extremely sensitive air-borne magnetometer for use in submarine detection and geological survey.[2] Some basic analytic expressions for the characteristics of the modulator were formulated at that time, but these are not sufficiently general to be useful in the over-all design of magnetic modulators.

In 1950 Williams and Noble[3] indicated that the fundamental limitation on signal level was the presence of Barkhausen noise inherent in the core material. This established the lowest detectable signal level at about 10^{-19} watt per cycle of bandwidth.

In 1951 Manley,[4] of the Bell Telephone Laboratories, set forth some expressions for output voltage, power gain, and rise time; but once again the results do not readily lend themselves to design procedures.

Thus the need for a general analysis of the operation of the magnetic modulator is felt. The analysis presented here is the result of a combined physical, geometrical, and mathematical interpretation of modulator operation. Experimental work has been carried out to verify the analytic expressions. On the basis of the analytic and experimental work some general design considerations are discussed.

Nomenclature

A = core cross section, square centimeters
a = voltage amplitude, volts
B = flux density, gauss
B_s = saturation flux density, gauss
E_2 = peak value of second-harmonic voltage, volts
e_0 = instantaneous value of output voltage, volts
f = frequency, cps (cycles per second)
F_1 = modulator figure of merit, $\sec^{-1}$ (second^{-1})
G = power gain, dimensionless
H = magnetizing force, oersteds
H_c = coercive force, oersteds
H_m = peak value of excitation magnetizing force, oersteds
H_0 = magnetizing force produced by signal, oersteds
H_s = saturation value of magnetizing force, oersteds
I_m = peak value of excitation current, amp (amperes)
I_0 = signal current, amp
i_e = instantaneous value of excitation current, amp
i_{sig} = instantaneous value of signal current, amp
L = signal winding inductance, henrys
l = mean core length, centimeters
N_0 = number of turns in output winding, turns
n = order of harmonic component, dimensionless
P_{in} = signal power input, watts
P_o = useful power output, watts
R_L = load resistance, ohms
R_w = signal winding resistance, ohms
T = period of excitation, sec
t = general time variable, sec
t_d = time delay due to hysteresis, sec
t_r = rise time of trapezoidal excitation, sec
α = slope of flux density, gauss per sec
β = phase shift due to signal, rad (radians)
δ = width of output voltage pulses, sec
μ_a = average permeability of core, gauss per oersted
μ_m = maximum permeability of core, gauss per oersted
τ = time constant, sec
ψ = slope of trapezoidal excitation, oersted per sec

Mathematical Analysis of Operation

PRINCIPLE OF OPERATION

In its most general form, the magnetic modulator consists of a single high-permeability toroid wound with separate windings for excitation, signal, and output. The device differs from an ordinary modulator in that the even harmonics of the excitation (carrier) are amplitude-modulated by the signal while the carrier itself is not. Any one of the even harmonics can be employed as the output or the summation of several or all even harmonics can be used.

If the waveform of the excitation current has zero-axis symmetry, then the resulting core flux (and hence the voltage induced in the output winding) con-

---

Paper 58-1277, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted August 1, 1955; made available for printing September 17, 1958.

E. J. KLETSKY, Syracuse University, Syracuse, N. Y., was formerly with the Massachusetts Institute of Technology, Cambridge, Mass., and the Technische Hogeschool te Delft, Delft, The Netherlands, as a recipient of the Anton Coenen Torchiana Fellowship for 1955–56.

This paper is based on a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering at the Massachusetts Institute of Technology (MIT). The research described was made possible through support extended MIT by the Office of Naval Research and the Navy Bureau of Ships under contract *N5ori-07876 NR-025-164*.

The author wishes to express his appreciation to Prof. Thomas F. Jones, Jr., for suggesting this problem, and for his guidance and criticism while the work was carried out. He also wishes to thank Profs. T. S. Gray and A. B. Van Rennes and the staff of the electronic nuclear instrumentation group of the MIT Servomechanisms Laboratory for their helpful suggestions and criticisms during the course of this work.

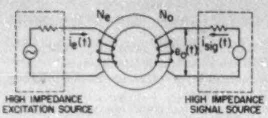

Ne
No
$i_e(t)$
$e_o(t)$
$i_{sig}(t)$
HIGH IMPEDANCE EXCITATION SOURCE
HIGH IMPEDANCE SIGNAL SOURCE

Fig. 1. High-impedance modulator

tains only fundamental and odd-harmonic components because of the odd-symmetrical character of the hysteresis loop of the core. If a signal is applied to the signal winding, even-harmonic components also appear in the output. For small values of signal, a linear relation exists between the signal and the amplitude of the even-harmonic components. The sense of the signal is preserved in the phase of the output with respect to the excitation.

The extremely low zero-drift of the modulator arises from the symmetrical character of the hysteresis loop around which the device operates. This symmetry is not disturbed by either temperature or frequency.

There are many modes of operation of the modulator. Excitation and/or signal may be supplied from voltage or current sources or from sources of finite impedance. The output may be taken as a voltage or a current. A single winding may serve two or more functions, thus allowing winding space to be used to best advantage.

The analysis which follows treats the case where both signal and excitation sources are of high impedance and the output is taken as a voltage.

NO-SIGNAL OPERATION

The circuit to be analyzed is shown in Fig. 1. Functionally the output voltage is given by

$$e_o(t) = -N_oA\frac{dB}{dt} = -N_oA\frac{d}{dt}\,[B(H)]$$

where

$$H = f(i_e) \tag{1}$$

In order to set up useful mathematical relations for the operation of the modulator, it is necessary to describe analytically the nonlinear flux-current characteristic. A piece-wise linear approximation[5] incorporating maximum permeability, zero saturation permeability, and neglecting hysteresis has been used.

Fig. 2 shows a geometrical interpretation of equation 1 in which this approximation to the flux-current characteristic is used. Sinusoidal excitation current can be approximated by the trapezoid shown for simplicity in the analysis. The validity of this approximation is borne out experimentally.

Fig. 2 indicates that if the excitation current is trapezoidal, the flux is also a trapezoidal function of time. The output voltage, found from the time derivative of the flux, consists of the sum of a set of positive pulses and a similar set of negative pulses shifted π rad. A Fourier series expansion shows that the output voltage is

$$e_o(t) = \sum_{n=1}^{\infty} \frac{2a}{n\pi}\,[1-(-1)^n]\,\sin\frac{n\pi\delta}{T}\,\cos\frac{2n\pi t}{T} \tag{2}$$

Note that $e_o(t)$ contains only odd harmonics, since when n is even, the expression goes to zero.

OPERATION WITH SIGNAL

When a signal I_o (corresponding to a magnetizing force equal to H_o) is applied, the output is altered as shown by the dotted lines in Fig. 2. With signal, the positive set of pulses is advanced by amount β rad and the negative set retarded by the same amount. Under these conditions, the output voltage is

$$e_o(t) = \sum_{n=1}^{\infty} \frac{2a}{n\pi}\sin\frac{n\pi\delta}{T}\left\{[1-(-1)^n]\cos n\beta \times \cos\frac{2n\pi t}{T} - [1+(-1)^n]\sin n\beta\,\sin\frac{2n\pi t}{T}\right\} \tag{3}$$

The geometrical parameters a, δ, and β of equation 3 can be replaced by electrical equivalents derived from the geometry of Fig. 2:

$$\tan\alpha = \mu_m \tan\psi = \mu_m\frac{H_m}{t_r} \tag{4}$$

$$\delta = 2\frac{H_s}{\tan\psi} = 2\frac{H_st_r}{H_m} \tag{5}$$

$$a = -N_oA\times 10^{-8}\frac{dB}{dt}$$

$$= -N_oA\mu_m\frac{H_m}{t_r}\times 10^{-8} \tag{6}$$

$$\beta = \frac{H_o}{\tan\psi} = \frac{t_rH_o}{H_m}(\text{sec}) = \frac{2\pi t_rH_o}{TH_m}\,(\text{rad}) \tag{7}$$

When the indicated substitutions are made and it is noted that under ordinary operating conditions $H_o << H_m$, the output voltage components become

$$e_o(t) = -\sum_{n\ \text{odd}}^{\infty} \frac{4N_oA\mu_mH_m}{n\pi t_r}\sin\left(\frac{2n\pi t_r}{T}\frac{H_s}{H_m}\right)\times \cos\frac{2n\pi t}{T} \tag{8}$$

$$e_o(t) = -\sum_{n\ \text{even}}^{\infty} 8N_oA\mu_mfH_o\times 10^{-8}\times \sin\left(\frac{2n\pi t_r}{T}\frac{H_s}{H_m}\right)\sin\frac{2n\pi t}{T} \tag{9}$$

Thus, with the assumption that the signal be small, equation 9 shows that the amplitudes of the even-harmonic components of the output voltage are linearly related to the signal.

In most applications, only the second-harmonic component is used as an output. This is done for two reasons:

1. Simple narrow-band amplification can be used.
2. The impedance level of the source for the second harmonic is lower than that for any higher harmonic.

For the second harmonic, equation 9 becomes

$$e_2(t) = \left[-3.2\pi N_o^2\frac{A}{l}\mu_mfI_o\times 10^{-8}\times \sin\left(4\pi f\frac{H_s}{H_m}t_r\right)\right]\sin 4\pi ft \tag{10}$$

VARIATION OF OUTPUT WITH AMPLITUDE OF EXCITATION

Equations 9 and 10 indicate that the output voltage is a function of the amplitude and rise time of the excitation as well as a function of the signal. Therefore, an optimum excitation signal is indicated. Its characteristics are found by setting

$$4\pi f\,(H_s/H_m)\,t_r = \pi/2$$

Thus,

$$\frac{H_m}{t_r} = 8H_sf = \tan\psi \tag{11}$$

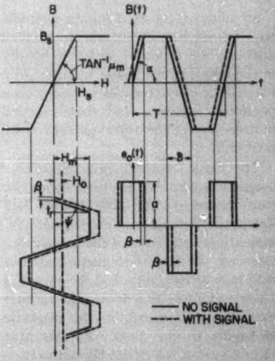


Fig. 2. Geometrical interpretation of modulator operation

Vol.
Pt.
195

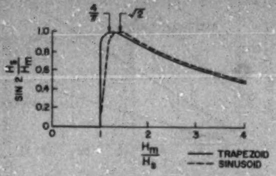


Fig. 3. Variation of output voltage with excitation amplitude

Equation 11 indicates the optimum slope of the excitation waveform in the unsaturated region.

If the excitation waveform is sinusoidal and given by

$$H(t)=H_m \sin 2\pi ft \tag{12}$$

then its initial slope is

$$\left.\frac{dH}{dt}\right|_{t=0}=2\pi fH_m \cos 2\pi ft\bigg|_{t=0}=2\pi fH_m = \tan\psi \tag{13}$$

If this value of slope is assumed to be constant in the unsaturated region, then the optimum amplitude of excitation can be found by equating equations 11 and 13.

$$\tan\psi=8H_sf=2\pi fH_m \quad \therefore \quad \frac{H_m}{H_s}=\frac{4}{\pi} \tag{14}$$

The variation of output voltage (for fixed signal) under the foregoing assumption is found by introducing equation 13 into equation 10. This yields, neglecting the minus sign,

$$\frac{E_2}{8N_oA\mu_mfH_o\times 10^{-8}}=\sin 2\frac{H_s}{H_m} \tag{15}$$

A normalized plot of this variation is shown in Fig. 3. The approximation to the magnetic characteristic accounts for the discontinuity in curve at $H_m/H_s=1$.

An analysis based directly on a sinusoidal excitation waveform[2] yields the dotted curve shown in Fig. 3.

TIME CONSTANT

The dynamic response of the modulator and its associated output filter is a function of the dynamic characteristics of both the signal-input winding and the output filter. The design of the output filter depends on the application of the modulator. Its dynamic characteristics normally can be synthesized to give any reasonable response. Since the output filter can be designed to give better dynamic response than the signal-input winding, the dynamic behavior of the modulator is determined by a single parameter; the time constant of the signal-input winding.

In so far as signal is concerned, the signal-input winding consists of an inductance L in series with the d-c winding resistance R_w. Thus the winding-time constant is given by

$$\tau=\frac{L}{R_w} \tag{16}$$

An average value of $u=\mu_a$, corresponding to the slope of the magnetization curve is used in the calculation of inductance, rather than the maximum value μ_m. The winding-time constant becomes

$$\tau=\frac{L}{R_w}=4\pi\frac{N_o^2\mu_aA}{lR_w}\times 10^{-9} \tag{17}$$

The time delay caused by hysteresis is important in a system using phase-sensitive demodulation. As a first approximation this delay is a constant depending only on the shape of the hysteresis loop and the rise time of the excitation; see Fig. 4. The amount of time delay t_d is given by

$$t_d=\frac{H_c}{\tan\psi}=\frac{H_c}{H_m}t_r \tag{18}$$

For the case of sinusoidal excitation,

$$\frac{H_m}{t_r}=2\pi fH_m \quad \therefore \quad \frac{t_d}{T}=\frac{1}{2\pi}\frac{H_c}{H_m} \tag{19}$$

POWER GAIN

It has been found experimentally that, for high-impedance excitation, a reasonable Thèvenin equivalent for the modulator as seen by the load consists of a voltage source in series with an inductor. The voltage source is equal to the open-circuit second-harmonic voltage calculated from equation 10. The reactance of the inductor is equal to the output-winding reactance evaluated at the second-harmonic frequency.

Elementary circuit theory indicates that maximum power transfer to the load occurs when R_L equals the magnitude of the source reactance. (It is assumed that $R_w<<2\omega L$.) Thus, the power delivered to R_L can be shown to be

$$P_o=3.2N_o^2\frac{A}{l}\frac{\mu_m^2}{\mu_a}I_of\times 10^{-8} \tag{20}$$

The signal power dissipated in the signal winding is

$$P_{in}=I_o^2R_w \tag{21}$$

Maximum power gain is therefore

$$G_{max}=\frac{P_o}{P_{in}}=\frac{8}{\pi}\frac{L}{R_w}\left(\frac{\mu_m}{\mu_a}\right)^2 f \tag{22}$$

If the signal winding is used for both signal and output, then the power-gain expression must be modified to include the signal power lost in R_L. The maximum power gain then will be

$$G_{max}=\frac{8}{\pi}\frac{L}{R_w+R_L}\left(\frac{\mu_m}{\mu_a}\right)^2 f \tag{23}$$

FIGURE OF MERIT

Equation 22 can be rewritten in a form that expresses a figure of merit, F_1, for modulator operation:

$$F_1=\frac{\text{Power gain}}{\text{Time constant}}=\frac{G}{\tau}=\frac{8}{\pi}\left(\frac{\mu_m}{\mu_a}\right)^2 f \tag{24}$$

While equation 24 does not account for hysteresis, core loss, or stray effects, it does indicate a fundamental design criterion for the modulator.

The factor μ_m/μ_a occurs, since the permeability upon which the output voltage depends differs from the value used to calculate nominal inductance. If experimental evidence indicates that this is actually the case, the factor μ_m/μ_a can be considered as a figure of merit for core materials. In any event, for a given core material, the modulator figure of merit, F_1, depends only on the frequency of operation.

PRACTICAL 2-CORE CURRENT-EXCITED MODULATOR

If two identical single-core modulators are arranged so that their excitation windings are in series and their output windings opposing, the combination doubles the even-harmonic output components and reduces the odd-harmonic output. The expressions governing the operation of the 2-core modulator are

$$\text{Voltage output}=e_o(t)=\left[-6.4\pi N_o^2\mu_mf\frac{A}{l}I_o\times 10^{-8}\times \sin 2\left(\frac{H_s}{H_m}\right)\right]\sin 4\pi ft \tag{25}$$

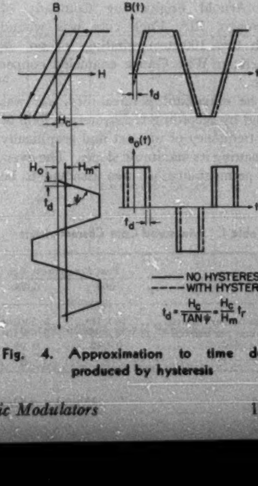


Fig. 4. Approximation to time delay produced by hysteresis

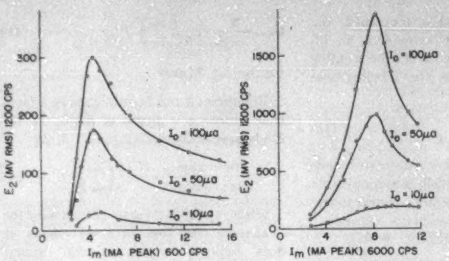


Fig. 5. Output voltage versus excitation amplitude

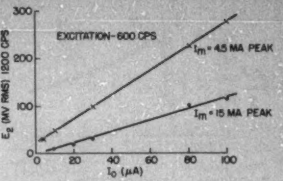


Fig. 6. Output voltage versus signal

$$\text{Time constant} = \tau = \frac{2L}{2R_w} = 4\pi\frac{N_o^2\mu_a A}{lR_w}\times10^{-9} \quad (26)$$

$$\text{Power output} = P_o = \frac{\left(\frac{2E_2}{\sqrt{2}}\right)^2}{2R_L}$$

$$= 6.4N_o^2\frac{A}{l}\frac{\mu_m^2}{\mu_a}I_o^2 f\times10^{-8} \quad (27)$$

$$\text{Power gain} = G_{\max} = \frac{P_o}{I_o^2(2R_w)} = \frac{8}{\pi}\frac{L}{R_w}\left(\frac{\mu_m}{\mu_a}\right)^2 f \quad (28)$$

$$\text{Figure of merit} = F_1 = \frac{8}{\pi}\left(\frac{\mu_m}{\mu_a}\right)^2 f \quad (29)$$

Comparison of equations 17 and 26, 23 and 28, and 24 and 29 indicates that the time constant, power gain, and figure of merit are the same for both 1-core and 2-core modulators. Thus, the only advantage of using a 2-core modulator is the suppression of odd-harmonic components.

Experimental Results

CORE PARAMETERS

The cores used in the experimental work were two 1-mil *4-79* Mo-Permalloy tape-wound toroids manufactured by the Arnold Engineering Company of Marengo, Ill.[6] Each core was wound with two 100-turn windings of no. *28* American Wire Gauge enameled copper wire.

The maximum permeability, μ_m, was found by recording the hysteresis loop at the frequency of interest and graphically measuring its maximum slope. The average permeability, μ_a, was assumed to be the slope of the normal magnetization curve. This normal curve was determined experimentally by plotting the location of the tips of the hysteresis loops as the excitation increased. A straight-line approximation was made and μ_a determined from the average slope in the unsaturated region.

These experimental techniques yield the results given in Table I.

Table I. Measured Core Characteristics

	Test Frequency, Cps	
	600	6,000
μ_m, gauss per oersted	10×10^4	5.3×10^4
μ_a, gauss per oersted	4×10^4	2.6×10^4
H_s, oersted	0.042	0.08
I_s, ma	4.0	7.6

OUTPUT-VOLTAGE CURVES

The 2-core current-excited modulator was chosen for experimental work because of its inherent rejection of odd-harmonics and its larger voltage output.

Experimental curves of output voltage versus excitation amplitude with fixed signal and output voltage versus signal with fixed excitation are shown in Figs. 5 and 6, respectively. Excitation frequencies of 600 and 6,000 cps were chosen to give reasonably large output signals and to be free of laboratory noise at the power frequency.

The curves shown in Fig. 5 can be normalized with respect to their peak amplitudes and also with respect to H_s. This is done for the case $I_o = 100$ μa (microamperes) and the results compared to the predicted theoretical behavior described by equation 15 in Fig. 7. If consideration is given to the difficulties in measuring the magnetic parameters of the cores and to the approximations made for the hysteresis loop, the expressions of the preceding "Mathematical Analysis of Operation" are in fair agreement with these experimental results.

The expression for output voltage in equation 25 becomes, upon substitution of the magnetic and physical parameters of the experimental cores,

$$e_o(t) = \left[-6.73\times10^{-5}\mu_m I_o f \sin 2\frac{H_s}{H_m}\right]\times$$

$$\sin 4\pi ft = E_2 \sin 4\pi ft \quad (30)$$

For operation at 600 cps, $\mu_m = 10\times10^4$ and $H_s = 0.042$ oersted [$I_s = 4.0$ ma (milliampere)]. Under these conditions the "transimpedance," the ratio of output voltage to input current, becomes

$$\frac{|E_2|}{I_o} = 6.73\times10^{-5}\times10\times10^4\times$$

$$600 \sin 2\left(\frac{4}{I_m}\right) = 4{,}040 \sin 2\left(\frac{4}{I_m}\right) \quad (31)$$

Equation 31 is an expression for the slope of the curve shown in Fig. 6. A comparison of these quantities is shown in Table II.

The effect of excitation frequency on transimpedance is shown in Fig. 8. At each experimental point, the amplitude of excitation current was adjusted to give maximum output voltage. Under these conditions, the transimpedance theoretically is given by

$$\frac{|E_2|}{I_o} = 6.73\times10^{-5}\mu_m f \quad (32)$$

When applying equation 32, it is important to note that μ_m is a frequency-dependent quantity. Shown in Fig. 8 are theoretical variations of transimpedance with frequency extrapolated from values of μ_m determined at 600 cps and 6,000 cps.

POWER-GAIN CURVES

Power-gain measurements, for signals of 100 μa, 30 μa, and 10 μa, were taken over five decades of resistance (R_L). The magnitude of the excitation current was adjusted to give maximum output voltage. Experimental results are indicated in Fig. 9.

The nature of the effective source impedance at the second-harmonic frequency can be found by comparing the shape of the normalized power-output curve with theoretical power-transfer curves for pure resistive and pure reactive

Table II. Transimpedance Comparison

I_m, Ma (600 Cps)	E_2/I_o (Equation 31), Ohms	E_2/I_o (Fig. 6), Ohms
4.5	3,940	3,620
15	2,040	1,830

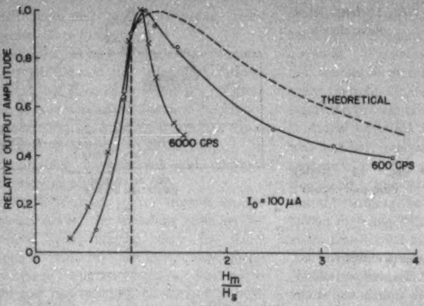


Fig. 7. Comparison of excitation amplitude curves

sources. If this comparison is made, the experimental curve is similar to the curve for reactive sources. Thus, a reasonable equivalent circuit for the modulator as seen by the load consists of a second-harmonic voltage source in series with a pure reactance.

Reference to the subsections on "Power Gain" and "Practical 2-Core Current-Excited Modulator" will indicate that maximum power gain should occur when the load resistance is equal to the output winding reactance calculated at the second-harmonic frequency. If the experimentally determined value of μ_a and the measured physical parameters are used, the inductance of the output winding is

$$L \text{ (at 600 cycles)} = \frac{N_o^2 \mu_a A}{l} \frac{4\pi}{10} \times 10^{-8}$$
$$= 168 \text{ millihenrys}$$

$$R_L \text{ (for maximum power gain)} = 4\omega L$$
$$= 2{,}540 \text{ ohms}$$

When the nominal bridge-measured value of $L = 75$ millihenrys is used,

$$R_L \text{ (for maximum power gain)} = 1{,}150 \text{ ohms}$$

Fig. 9 indicates a value of $R_L = 1{,}500$ ohms, which is bracketed by the calculated values.

The value of maximum power gain is given by equation 28. When the measured values of the parameters are used, the theoretical power gain becomes

$G_{\max} = 2{,}220$ (using calculated value of L)

$G_{\max} = 990$ (using nominal measured value of L)

Fig. 9 indicates a value of $G_{\max} = 1{,}400$, which is bracketed by the calculated values.

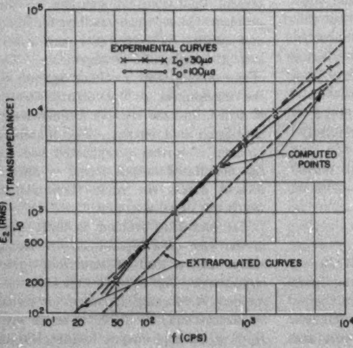
Conclusions and Design Considerations

Several aspects of the mathematical analysis of operation developed in a previous section have been verified by the work shown in the section on "Experimental Results." The correlation between the analysis and the experiment is sufficiently high to allow the analysis to be used as a basis for design of second-harmonic magnetic modulators. Presented in the following are conclusions and design criteria based on the work of these sections and also the work of other experimenters in the field.[3,7,8]

Choice of Modulator Configuration

Perhaps the simplest configuration consists of a single-core wound with a sufficient number of windings to supply excitation, signal, and output requirements. The major disadvantage of such a configuration is the presence of large odd-harmonic components in the output winding. These components are usually much larger than the desired second-harmonic output signal and, unless selective circuits are used, there exists the possibility of overloading the stage following the modulator, and generating spurious signals.

A much better configuration consists of two cores arranged so that the odd-harmonic components in each output winding cancel each other and hence do not appear in the output. The degree of odd-harmonic suppression depends on how well the magnetic characteristics of the two cores are matched. Normally two physically similar cores will be sufficiently well matched to give on the order of a 10-to-1 suppression of odd harmonics. This can be improved by selecting cores or by adjusting the windings on each core to give the effect of matched cores. Differences in the width of the hysteresis loop can be minimized by altering the number of excitation turns in one or the other of the cores. Similarly, differences in maximum flux density can be minimized by altering output-winding turns.



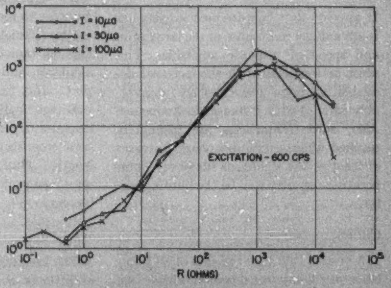


Fig. 8 (left). Transimpedance versus frequency

Fig. 9 (above). Power gain versus load resistance

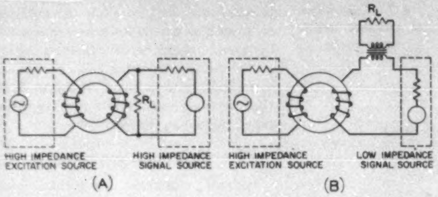

R_L
R_L
R_L
HIGH IMPEDANCE EXCITATION SOURCE
HIGH IMPEDANCE SIGNAL SOURCE
(A)
HIGH IMPEDANCE EXCITATION SOURCE
LOW IMPEDANCE SIGNAL SOURCE
(B)

Fig. 10 (left). Modulator circuits

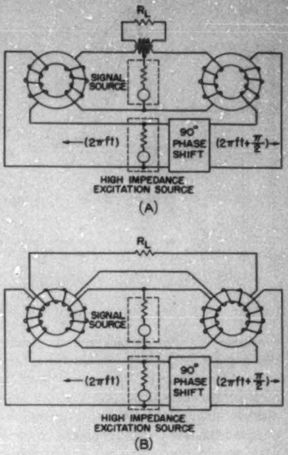


Fig. 11 (right). Push-pull circuits

Choice of Excitation

A most important consideration in the choice of an excitation source is that it be able to supply sufficient magnetomotive force to saturate the core. In addition, the waveform of the excitation must be free from the harmonic component used as the output. To insure this, adequate filtering in the excitation circuit is necessary. Modulator transimpedance is a function of excitation amplitude; thus amplitude stability is important to prevent changes in gain.

Choice of excitation frequency depends on many factors. Output signal frequency and amplitude are directly dependent on excitation frequency; hence equipment following the modulator may dictate the choice. The most important factors to be considered are given in the following.

Factors favoring low excitation frequency:

1. Economy of excitation power.
2. Small core loss, and hence less heating.
3. Low impedance levels.
4. Freedom from winding resonances.

Factors favoring high excitation frequency:

1. Large gain.
2. Fast time response thereby accommodating wider bandwidths of input signal.

It appears that high-frequency operation is advantageous because of the large gains and speed of response available. This must be tempered with the fact that desirable magnetic properties are unfavorably altered with increases in frequency. Also, since frequencies are ultimately reached where excitation power increases at least as the square of frequency, temperature rise may well be the limiting factor.

Choice of Output

A system that uses the sum of all even-harmonic components would utilize the maximum energy available as output. Such a system would require wide-band amplification and demodulation schemes.

Single-frequency output eliminates the foregoing difficulties. Experimentally, it has been found that the fourth- and sixth-harmonic amplitudes are smaller than the second-harmonic amplitude for a given signal. This, coupled with the fact that the second-harmonic impedance is lower than that of the higher harmonics, tends to indicate that the second-harmonic component is most desirable as the output frequency.

Choice of Core Material

Materials having high maximum permeability are best suited for use in magnetic modulators. Low core loss is important since it is desirable to operate the core at high frequencies without excessive heating. Thus a figure of merit for core materials operated at the upper end of their useful range can be defined as $\mu_m f_c$, where f_c is the frequency at which the core loss becomes appreciable.

Reference to the subsection, "Figure of Merit," indicates that an additional figure of merit for core materials might be μ_m/μ_a. Lack of time, however, has prevented experimental verification of its validity.

Core Shape and Size

Since very high permeabilities are sought, air gaps must be eliminated. This suggests the use of toroidal cores. In order to reduce core losses, lamination (or tape) thickness should be minimized, One-mil tape is common, 1/8-mil tape is available, and thinner tapes are in prospect. If a fixed volume is available for windings, both power output and time response depend directly on the ratio of core cross-sectional area to mean core length. Thus, the core size will depend on a compromise between the conflicting factors.

Windings

The excitation winding should take up as little of the available winding space as possible. Its wire size and number of turns will depend on the magnetomotive force required to saturate the core and the internal impedance of the excitation source. The winding should be distributed over the core so that the flux density is uniform across the area.

Separate windings for signal and output result in isolation between signal and output circuits but reduce sensitivity. Separate windings also allow impedance matching between signal source and modulator and between modulator and load. The main advantage of a single winding used for both signal and output is the increase in sensitivity.

Circuits for Realizing Maximum Power Gains

The use of a single winding for both signal and output is not always compatible with impedance matching for maximum power transfer. Three techniques present themselves as solutions to this problem: (1) separate windings for signal and output; (2) the use of a transformer; and (3) "push-pull" arrangements. The choice among these methods depends on the impedances of the sources to be matched, and the relative frequencies of the signal and output. The matching problem is further complicated because the modulated output signal is always present across the same terminals to which the signal is applied.

The matching problem is fairly easy in the case of a high-impedance signal source and a load of somewhat lower impedance. Fig. 10(A) indicates one method. A blocking capacitor for signal frequencies may be placed in series with R_L if signal and output frequencies are well separated and if winding resonances do not occur.

In Figs. 10 and 11 single-core modulators are shown for simplicity. However, the circuits are equally effective for 2-core modulators which (as shown under "Practical 2-Core Current-Excited Modulator") offer the advantage of suppressing fundamental and odd-harmonic components.

As the signal source impedance becomes lower, the matching problem becomes more difficult, since the signal source tends to short-circuit the output winding. The use of a series load resistance is not effective because of the undesirable loss of signal power in R_L. The circuit of Fig. 10(B) employs a transformer, which is used to present low-impedance to signal frequencies and high impedance at the output frequency. The major difficulty of this circuit lies in the realization that the transformer itself may act as a modulator and introduce output components bearing no relation to the input signal.

The push-pull circuit shown in Fig. 11(A) eliminates the nonlinear difficulties of Fig. 10(B) by virtue of the inherent even-harmonic suppression of such a system. It is necessary that the excitation sources be 90 degrees out of phase in order that the circuit act in a push-pull manner at the second-harmonic frequency. It should be noted that two reasonably well-matched magnetic modulators are necessary for this circuit to operate effectively. In addition, the dynamic characteristics of the center-tapped transformer must meet the same requirements as the transformer in Fig. 10(B).

A modified push-pull circuit which eliminates the need for an external transformer is shown in Fig. 11(B). Here the cores themselves act as the center-tapped transformer. Output is taken from a third set of windings on the core. Since these windings are isolated from the signal input windings the output-winding impedance does not affect the signal power delivered. By adjusting turns and wire size, the output-winding impedance can be independently matched to the load R_L. This circuit still requires that the excitation sources be 90 degrees out of phase.

References

1. Magnetic Modulators, E. P. Felch, V. E. Legg, F. G. Merrill. *Electronics*, New York. N. Y., vol. 25, no. 2, Feb. 1952, pp. 113–17.

2. Airborne Magnetometers for Search and Survey, E. P. Felch, W. J. Means, T. Slonczewski, L. G. Parratt, L. H. Rumbaugh, A. J. Tickner. *AIEE Transactions*, vol. 66, 1947, pp. 640–50.

3. Magnetic Circuits and Transformers (book), Electrical Engineering Staff, Massachusetts Institute of Technology. John Wiley & Sons, Inc., New York, N. Y., 1943.

4. Some General Properties of Magnetic Amplifiers, J. M. Manley. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 39, Mar. 1951, pp. 242–51.

5. Mathematical Analysis of Non-Linear Circuits, A. Boyajian. *General Electric Review*, Schenectady, N. Y., vol. 34, 1931.

6. Technical Data on Arnold Tape-Wound Cores. *Bulletin TC-101A*, The Arnold Engineering Company, Marengo, Ill., Mar. 1953.

7. The Fundamental Limitations of the Second-Harmonic Type of Magnetic Modulator as Applied to the Amplification of Small D.C. Signals, F. C. Williams, S. W. Noble. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, no. 58, Aug. 1950, pp. 445–59.

8. The Design of a Practical D.C. Amplifier Based on the Second-Harmonic Type of Magnetic Modulator, S. W. Noble, P. J. Baxandall. *Ibid.*, vol. 99, no. 70, Aug. 1952.

9. A Simple Magnetic Modulator for Conversion of Millivolt D-C Signals, Gunnar Wennerberg. *Electrical Engineering*, vol. 70, Feb. 1951, pp. 144–47.

On Feedback in Magnetic Amplifiers

Part I—Single Feedbacks

L. A. FINZI
MEMBER AIEE

J. J. SUOZZI
ASSOCIATE MEMBER AIEE

FEEDBACK TECHNIQUES are receiving considerable attention in the magnetic-amplifier field and many ways have been suggested in which the output quantity can be chosen and employed to modify the controlling action of the signal. This paper examines and compares a variety of feedback configurations that are associated with 2-core full-wave self-saturating amplifiers. A companion paper, Part II, extends the study to combined feedbacks, as used in "self-balancing" amplifiers.[1]

Expressions are presented for input and output impedances, and gains in the steady state with the loading effects of the components of the feedback loop taken into account where necessary. The transient and frequency response are treated, extending familiar "slow-transient" approaches to the various circuitries and showing the influence of coupled feedback circuits upon the gain-bandwidth product. A difference-equation approach is used subsequently and "modular" time constants are derived which befit the jump-function character of the variables.

The present study is limited to resistive loads and feedback networks without frequency-dependent elements; the operation over the "negative branch" of the transfer characteristic is not considered.

I. Terminology: Algebraic Conventions

Since a unified language is lacking in the literature, definitions are given as to the output quantity chosen to perform feedback and the manner in which this quantity is used. The terms "load-voltage feedback," "load-current feedback" and "gate-winding current feedback," with or without potentiometers and shunts, refer to the quantity chosen and are self-explanatory.

"Electric" (or galvanic) feedback exists when the output quantity or some fraction thereof is applied to the signal circuit. It may act in series with the signal source ("series" electric feedback), or across the terminals of the signal winding ("shunt" electric feedback).

"Magnetic" or mmf (magnetomotive force) feedback occurs when the output quantity acts upon the resetting core through windings other than the signal winding. A variety of practical situations are covered by this definition.

Mmf of windings on any core are positive when in the direction of positive fluxes, and positive currents give positive

---

Paper 58-1231, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing August 26, 1958.

L. A. Finzi is with Carnegie Institute of Technology, Pittsburgh, Pa., and J. J. Suozzi is with the Bell Telephone Laboratories, Inc., Whippany, N. J.

This work was performed at the Carnegie Institute of Technology under the sponsorship of the Minneapolis-Honeywell Regulator Company, Brown Instruments Division, Philadelphia, Pa. Portions of the paper are abstracted from a dissertation by J. J. Suozzi, submitted to the Carnegie Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Useful suggestions by Prof. A. B. Heins and A. G. Milnes of Carnegie Institute of Technology and by G. F. Pittman, Jr., of the Westinghouse Electric Corporation are gratefully acknowledged.

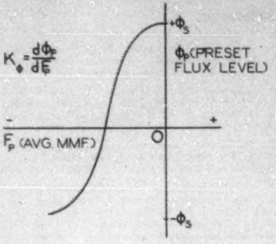


Fig. 1. Core-control curve

mmf when entering the dot terminal of a winding. The "positive branch" of a transfer characteristic covers the common type of operation in which the flux of the gating core rises from the preset flux level, ϕ_p, to saturation $+\phi_s$, while the flux of the other core during the same half-cycle resets from positive saturation. The amount of reset, $\phi_s-\phi_p$, increases for algebraic decreases of the half-cyclic average mmf F_p, of all the currents acting upon the resetting core. Hence larger outputs result from algebraic increases of signal, regardless of bias or feedback. Positive feedbacks act in the same direction as positive increases of signal.

Given a pair of cores with gate windings N_g, and signal windings N_c, any particular type of amplifier circuitry could be chosen as the starting point for further modifications by feedback. At earlier times, the simple saturable reactor has been used for this. In the following the self-saturating amplifier (bridge, center tap, or doubler) is preferred; more specifically the bridge circuit is taken as the basis of reference, although the center-tap circuit would have been equally suitable. Systems derived from the doubler, or from an equivalent saturable reactor with "100% extrinsic feedback," are more vulnerable to unblocking effects, which limit the validity of the treatment.

II. The Control Curve

It is assumed that the initial flux ϕ_p, of the gating core can be related simply to the half-cyclic average F_p, of all mmf that acted upon it in the preceding reset half-cycled by means of a "core-control curve" $\phi_p=f(F_p)$; see Fig. 1. It is also assumed that the resetting process initiates when the alternating voltage at the gate terminals passes through zero.

The latter assumption is true if the signal voltage is unidirectional and has the same wave form and phase before rectification as the gate voltage. It remains acceptable practically in more general situations (e.g., sinusoidal gate voltages and d-c signals) for self-saturating amplifiers even when comparatively large negative feedbacks are introduced. The assumed uniqueness of a control curve is more questionable. The general understanding of dynamic core behavior, incomplete as it may be, leads one to doubt that different half-cyclic processes arriving at a certain ϕ_p with different instantaneous rates should result ultimately in a common value of the half-cyclic average F_p, but the acceptance of a core-control curve remains the best available for the purposes at hand and is preferable by far to arbitrary postulations of idealized cores. Core-control curves obtained from transfer characteristics measured at the pertinent gate voltage and frequency, preferably with low control circuit constraint, proved quite adequate in the experimentation. Departures from analytical predictions arose rather for cases in which the rectifiers blocked or unblocked at improper times.

III. Steady-State Gains, Input and Output Resistances

III-1. DEFINITIONS

Capital letters indicate half-cyclic average values. Instantaneous values are indicated by lower-case letters.

E_g = a-c supply voltage. (For $e_g=E_{gm}\sin\omega t$, $E_g=(2/\pi)\,E_{gm}$; not $E_{gm}/\sqrt{2}$)
E = gross output voltage (i.e., the part of the supply voltage not absorbed by the half-cyclic swing of fluxes)
V_L = voltage across the load resistor R_L. (I_L = load current $=V_L/R_L$)
E_c = applied signal voltage (not necessarily a pure direct current)
I_i = input current
I_c = current *in the signal windings N_c. ($I_c=I_i$ except for circuits with shunt-electric feedback)
R_i = input resistance $=dE_c/dI_i$
$K_E=dE/dE_c$ = gross output-voltage gain, static
$K_{EN}=K_EN_c/N_g$ = gross output-voltage gain per turn, static
$K_V=dV_L/dE_c$ = net output-voltage gain (load-voltage gain), static
$K_A=dI_L/dI_i$ = current gain, static
$K_{VA}=K_V\times K_A$ = load volt-ampere gain, static
$K_{EA}=K_E\times K_A$ = gross volt-ampere gain, static

III-2. BASIC SELF-SATURATING CIRCUITS

The familiar bridge circuit and also the center-tap and doubler circuits are presented in Appendix I, mainly to explain certain definitions of symbols. The resistance of a gate winding is indicated by r_g, and r_{gr} is the sum of r_g and of the linearized forward resistances $r\gamma$ (and

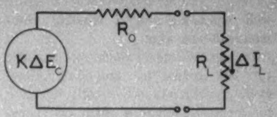


Fig. 2. Thevenin equivalent of output circuit

also $r\sigma$ for the bridge) of the conducting diodes in the gating path. Rectifier leakage is assumed negligible and may be considered separately as a particular kind of negative magnetic feedback; see section III-4(d). The differential equation of the gating circuit during positive half-cycles, with $\dot\phi=d\phi/dt$ is

$$e_g=N_g\dot\phi_1+(R_L+r_{gr})i_{g1}=N_g\dot\phi_1+R_gi_L \qquad (1)$$

If steady state with constant signal prevails, a half-cyclic time integration yields

$$E_g=2fN_g(\phi_s-\phi_p)+R_gI_L$$
$$=2fN_g(\phi_s-\phi_p)+E \qquad (2)$$

(To include cases of "subnormal" excitation, i.e., $E_g<4fN_g\phi_s$, the restriction $2fN_g(\phi_s-\phi_p)<E_g$ is added to specify the range of ϕ_p for which equation 2 is valid. In other words, the initial ϕ_p must be such that $+\phi_s$ is reached within the half-cycle.)

The control circuit resistance R_c is the sum of the resistances $2r_c$, of the control winding and, in general, of an additional resistor R_k in series. With $R_c=R_k+2r_c$, the differential equation of the signal circuit is

$$e_c=R_ci_c+N_c(\dot\phi_1+\dot\phi_2) \qquad (3)$$

whence

$$E_c=R_cI_c \qquad (4)$$

if the steady state is considered. The gross output-voltage gain of the basic amplifiers (marked by an asterisk in the following) is

$$K_E{}^*=\frac{dE}{dE_c}=\frac{dE}{d\phi_p}\,\frac{d\phi_p}{dF_p}\,\frac{dF_p}{dE_c}=2fN_gK_\phi\frac{N_c}{R_c} \qquad (5)$$

where $K_\phi=d\phi_p/dF_p$ is the slope of the core-control curve, Fig. 1, at any ϕ_p. The corresponding gross output-voltage gain per turn is

$$K_{EN}{}^*=K_E{}^*\frac{N_c}{N_g}=2fK_\phi\frac{N_c{}^2}{R_c} \qquad (6)$$

III-3. MODIFICATIONS BY FEEDBACK: PROCEDURE AND PRESENTATION OF RESULTS

Modifications of the original bridge-amplifier by feedbacks of various kinds are analyzed from the equations of the network in terms of half-cyclic average voltages and currents, and from the mmf of reset F_p, as the sum of the contributions given by the various windings on the core; the algebra is straightforward and does not need to be given in detail here.[2] Care has been taken to preserve loading terms (i.e., terms containing shunt and potentiometer resistances, etc.) in the expressions obtained. Loading effects of the feedback networks of magnetic amplifiers are less likely to be negligible than in similar configurations with electronic amplifiers. (Certain simplifications of some final expressions of the analysis are stated at pertinent places in the tabulation of results.)

The quantities given in Appendixes I through III are the gross voltage gain, K_E; the input resistance, R_i; the output resistance R_o; and the generator gain $\mathcal{K}$, of the Thevenin equivalent of the output circuit of the amplifier system; see Fig. 2. For all cases, the load voltage gain K_V and the current gain K_A are given by

$$K_V = \mathcal{K}R_L/(R_L+R_o)$$

and

$$K_A = \mathcal{K}R_i/(R_L+R_o)$$

The winding polarities of the schematic diagrams and the upper algebraic signs of the expressions apply for negative feedback action; the lower algebraic signs refer to positive feedback action. N_b and R_b are turns and resistances of bias circuits, not shown on the schematics. (Other quantities presented in the tables refer to the study of transient and frequency response; their meaning is explained in section IV.)

The various classes of magnetic and electric feedback considered are described briefly in the following.

III-4. MAGNETIC FEEDBACKS

(a). *Load-Current Feedback into Separate N_f-Coils, Appendix II(A):* The unidirectional load current, or a part thereof, derived from a shunt r_s, in series with the load R_L, is fed into two N_f-coils, thus providing an additional negative (or positive) mmf.

(b). *Load-Voltage Feedback into Separate N_f-Coils, Appendix II(B):* Here the N_f-coils are supplied from the unidirectional load voltage across the load or a part thereof, obtained from a potentiometer R_1+R_2 across the load.

(c). *Gate-Winding-Current Feedback in Cross-Connected Feedback Windings, Appendix II(C):* Here, the feedback turns N_f on one core are connected in series with the gate winding of the opposite core, so that the gate current of the firing core gives a contribution (negative or positive), to the mmf of the resetting core.

(d). *Negative Feedback Action of Shunted Gate Rectifiers, Appendix II(D), (E), and (F):* The self-saturating rectifiers γ of the bridge, center tap, and doubler circuits, are shunted by resistors R_{sh}, allowing a current to flow in the gate winding of the resetting core with a negative contribution to the mmf of the resetting core. The reverse rectifier voltages are different in the three basic circuits and accordingly each circuit is treated separately. For bridge and center-tap circuits, the feedback resembles a negative magnetic load-voltage feedback of item (b). For the doubler, the feedback is more nearly a negative current magnetic feedback such as in item (a). The shunting of rectifiers with adequately high R_{sh} resistors offers a simple way to reduce the voltage gain, if so desired. The adverse effects of rectifier leakage on voltage gains are also expressed by this treatment if the "reverse resistance" of the leaky rectifier is approximated by the linearized constant value R_{sh}.

(e). *The Positive Feedback Action of Unblocking Rectifiers.* The treatment given in this paper assumes that no unblocking of the rectifiers occurs, but a passing remark may be appropriate. Phenomena that may conceivably cause the rectifier of the resetting core to conduct in its forward direction result in a positive contribution to the average mmf of the resetting core. This is a sort of spurious positive feedback, opposite in its essence to the negative feedback of rectifier leakage. If the amplifier in the absence of unblocking has a rather high gain K_E to start with, any agent that induces unblocking (i.e., unusual constraints in the signal crcuit or, in addition, some inductance in the gate circuit) may result in infinite or negative gains over some range of the characteristic, and thus in unwelcome instabilities.

(f). *The Simple SR (Saturable Reactor).* The "inherent" negative feedback of the SR is a type of negative magnetic feedback. In the series-connected SR, the output current flowing in the gate winding of the resetting core gives a negative contribution to the mmf F_p. The parallel connected SR is a limiting case of the doubler with shunted gate rectifiers.

III-5. ELECTRIC FEEDBACKS

(a). *Series Load-Current Feedback, Appendix III(A):* The voltage drop across a shunt resistor r_s, in series with the load R_L, is injected into the signal circuit in series with the signal source.

(b). *Series Load-Voltage Feedback, Appendix III(B):* The voltage obtained from a potentiometer across the load is injected into the signal circuit in series with the signal source. Here (and in subsequent figures in the appendixes) certain approximations, as stated, have been introduced in view of the usual orders of magnitude of certain terms in the relationship.

(c). *Shunt Load-Current Feedback, Appendix III(C):* The voltage drop as in item (a) is applied across the terminals of the signal winding through a resistance R_a. The resistance $2r_c$ of the signal windings is expressed explicitly in the analysis, i.e., $(R_c = R_k + 2r_c)$.

(d). *Shunt Load-Voltage Feedback, Appendix III(D):* The potentiometer voltage is applied across the terminals of the signal winding through a resistance R_a.

IV. Transient and Frequency Response

IV-1. BASIC AMPLIFIER AND ITS DIFFERENCE EQUATIONS

If the amplifier processes do not repeat half-cycle by half-cycle, the half-cyclic time integration of equation 1 performed from

$$t = \frac{(n-1)\pi}{\omega} \text{ to } t = \frac{n\pi}{\omega}$$

with n = a positive integer, yields more generally

$$E_{g(n)} = 2fN_g(\phi_s - \phi_{p(n-1)}) + E_{(n)} \qquad (7)$$

Equation 7 relates the gross output average voltage $E_{(n)}$ of the nth half-cycle to the corresponding half-cyclic average gate voltage $E_{g(n)}$, and to the preset flux level $\phi_{p(n-1)}$, i.e., to the flux level reached by the resetting core at end of the previous half-cycle.

Equation 3 integrated over the nth half-cycle yields

$$E_{c(n)} = R_cI_{c(n)} + 2fN_c(\phi_{p(n)} - \phi_{p(n-1)}) \qquad (8)$$

Equation 8 relates the half-cyclic average signal voltage and current $E_{c(n)}$ and $I_{c(n)}$ and the difference of the preset flux levels $\phi_{p(n)}$ and $\phi_{p(n-1)}$ attained at the end of the nth and $n-1$th half-cycle respectively.

In section II, the flux level ϕ_p, attained by a core at the end of its half-cycle of reset, has been related to the half-cyclic average F_p of all the mmf that acted upon the core during that half-cycle by means of a given core control curve $\phi_p = f(F_p)$. No additional inaccuracies are introduced if the same curve is used in the transient analysis to relate $\phi_{p(n)}$ to $F_{p(n)}$ at least as long as the assumption of negligible unblocking holds in the transient condition.

IV-2. THE USE OF APPROXIMATE DIFFERENTIAL EQUATIONS

A time-honored approach is that in which equation 8 is converted into an approximate differential equation:

$$E_c = R_c I_c + N_c d\phi_p/dt \tag{9}$$

where the discrete values of equation 8 are substituted by continuous variables.

In the basic amplifier without bias windings $F_p = N_c I_c$, and thus equation 9 is converted into the approximate differential equation

$$E_c(t) = \frac{R_c}{N_c} F_{p(t)} + N_c \frac{d\phi_{p(t)}}{dt}$$
$$= \frac{R_c}{N_c} F_{p(t)} + N_c K_\phi \frac{dF_{p(t)}}{dt} \tag{10}$$

If K_ϕ is a constant (i.e., if the voltage gain $K_E{}^* = 2fN_g K_\psi N_c/R_c$ does not vary significantly over the range of outputs of interest), the transient of $F_{p(t)}$, (and thus of $\phi_{p(t)}$), is characterized by a simple time constant

$$\tau_a = K_\phi \frac{N_c{}^2}{R_c} = \frac{1}{2f} K_E{}^* \frac{N_c}{N_g} = \frac{1}{2f} K_{EN}{}^* \text{ seconds} \tag{11}$$

(If $K_E{}^*$ varies significantly over the range, graphical integrations of equation 9 remain possible.)

The transient output $E_{(t)}$ is obtained then, also as a continuous time function using $\phi_{p(t)}$ in equation 2. This approach is adequate for transients that are long compared to the half-cycle of the gate frequency. The concept of a half-cycle "transportation lag" has been suggested to account for the fact that the output of one half-cycle depends on the preset flux level found at the end of the previous half-cycle. However, the substitution of an approximate continuous differential equation for the original equation 8 is, in itself, a source of error of the same order of magnitude as that attempted correction. For fast transients, where this matter becomes significant, a consistent solution of the difference equations is preferable; see section IV-5.

As is generally known, if additional circuits are coupled directly to the control circuit, such as a bias circuit with N_b turns per core and total resistance R_b, equation 10 modifies into

$$\tau_a = \frac{1}{2f} K_{EN}{}^* \left(1 + \frac{R_c}{R_b} \frac{N_b{}^2}{N_c{}^2}\right) \text{ seconds} \tag{12}$$

Here, τ_a is still the time constant of the transient of both $F_{p(t)}$ and $\phi_{p(t)}$.

IV-3. EXTENSIONS OF APPROXIMATE APPROACH TO AMPLIFIERS WITH FEEDBACK

One may extend the slow-transient approach to amplifiers with magnetic or electric feedbacks, writing the instantaneous differential equations of the various loops, converting them into difference equations, and then approximating their discrete variables with continuous time functions. The procedure is illustrated in two examples.

For magnetic load-current feedback, Appendix II(A), the control circuit equation remains equation 9. Then $I_{c(t)}$ is expressed in terms of $F_{p(t)}$ as

$$F_{p(t)} = N_c I_{c(t)} \mp N_f I_{f(t)} \mp N_b I_{b(t)} \tag{13}$$

where the upper signs apply for negative feedback (and for negative bias respectively). The voltage equation of the bias circuit, in approximate continuous form is

$$E_{b(t)} = R_b I_{b(t)} \mp N_b \frac{d\phi_{p(t)}}{dt} \tag{14}$$

The voltage equation of the gate and feedback circuit in difference form is

$$E_{g(n)} = 2fN_g(\phi_s - \phi_{p(n-1)}) + \frac{R_f R_g{}'}{r_s} I_{f(n)} \mp$$
$$2fN_f \frac{R_g}{r_s} (\phi_{(pn)} - \phi_{p(n-1)}) \tag{15}$$

from which is obtained the approximate differential equation

$$E_{g(t)} = 2fN_g(\phi_s - \phi_{p(t)}) + \frac{R_f R_g{}'}{r_s} I_{f(t)} \mp$$
$$\frac{R_g}{r_s} N_f \frac{d\phi_{p(t)}}{dt} \tag{16}$$

$F_{p(t)}$ is expressed by $F_{p(t)} = (\phi_{p(t)} - \phi_K)/K_\phi$ if the control curve is linearized over the range of interest, where ϕ_K is the intercept of the straight line with the ϕ_p axis and $\phi_{p(t)}$ is expressed as a function of time by the differential equation

$$E_c \mp E_b \frac{N_b}{N_c} \frac{R_c}{R_b} \mp \beta(E_g - 2fN_g\phi_s)$$
$$= (\phi_{p(t)} - \phi_K)\left(\frac{1 \pm \beta K_E{}^*}{K_\phi N_c/R_c}\right) +$$
$$N_c\left(1 + \frac{N_f{}^2 R_c R_g}{N_c{}^2 R_f R_g{}'} + \frac{N_b{}^2 R_c}{N_c{}^2 R_b}\right)\frac{d\phi_{p(t)}}{dt} \tag{17}$$

Hence, for constant E_c, E_b, and E_g, the transient of $\phi_{p(t)}$ is given by

$$\phi_{p(t)} - \phi_K$$
$$= \left[\frac{E_c \mp E_b \frac{N_b}{R_b} \frac{R_c}{N_c} \mp \beta(E_g - 2fN_g\phi_s)}{1 \pm \beta K_E{}^*}\right]$$
$$K_\phi N_c/R_c \left(1 - \epsilon^{-\frac{t}{\tau_a}}\right) \tag{18}$$

where

$$\tau_a = \frac{1}{2f} K_{EN}\left(1 + \frac{N_f{}^2 R_c R_g}{N_c{}^2 R_f R_g{}'} + \frac{N_b{}^2 R_c}{N_c{}^2 R_b}\right) \tag{19}$$

with the symbols defined in Appendix II(A). It is seen that the feedback loop introduces a coupling term similar to that of a bias circuit. Writing the voltage equation of the feedback loop yields

$$r_s I_{L(t)} = (R_f + r_s) I_{f(t)} \mp N_f d\phi_{p(t)}/dt$$

and, substituting in equation 16, the load current $I_{L(t)}$ is obtained as

$$I_{L(t)} = \left(\frac{r_s + R_f}{R_f R_g{}'}\right)(E_g - 2fN_g(\phi_s - \phi_{p(t)})) \pm$$
$$\frac{N_f}{R_g{}' R_f}\left(R_g - \frac{R_g{}' R_f}{r_s}\right)\frac{d\phi_{p(t)}}{dt} \tag{20}$$

According to this approach, $I_{L(t)}$ is the sum of two exponential terms, both with the same time constant τ_a; the second right-hand term expresses the fact that transient voltages are induced in the feedback loop associated with the output. The ambiguous concept of a continuous $I_{L(t)}$ does not warrant pursuing this point further.

For series electric load-current feedback, the difference equation of the control circuit is

$$E_{c(n)} = (R_c + r_s) I_{c(n)} \pm r_s I_{L(n)} +$$
$$2fN_c(\phi_{p(n)} - \phi_{p(n-1)}) \tag{21}$$

The difference equation of the gate circuit is

$$E_{g(n)} = R_g I_{L(n)} \pm r_s I_{c(n)} + 2fN_g(\phi_s - \phi_{p(n-1)}) \tag{22}$$

Equations 21 and 22 for E_c constant become in approximate form

$$E_c = (R_c + r_s) I_{c(t)} \pm r_s I_{L(t)} + N_c d\phi_{p(t)}/dt \tag{23}$$

$$E_{g(t)} = R_g I_{L(t)} + 2fN_g(\phi_s - \phi_{p(t)}) \pm r_s I_{c(t)} \tag{24}$$

With $F_{p(t)} = N_c I_{c(t)} \mp N_b I_{b(t)}$ and with constant E_g, the general equation for the control circuit becomes

$$E_c \mp E_b \frac{N_b}{N_c} \frac{R_c}{R_b} (1 + \nu) \mp \beta(E_g - 2fN_g\phi_s)$$
$$= (\phi_{p(t)} - \phi_K)\left(\frac{1 + \nu \pm \beta K_E{}^*}{K_\phi N_c/R_c}\right) +$$
$$N_c\left(1 + \frac{N_b{}^2}{N_c{}^2} \frac{R_c(1 + \nu)}{R_b}\right)\frac{d\phi_{p(t)}}{dt} \tag{25}$$

This approximate differential equation describes the transient of $\phi_{p(t)}$. The time constant τ_a, is given in Appendix III(A). The output current $I_{L(t)}$ in this case is affected by the control current through the galvanic coupling, but once more its time constant is τ_a.

All other types of feedback may be treated in the same manner. The corresponding expression of τ_a are presented in the Appendixes. The effect of coupled loops arising from the introduction of magnetic (or shunt electric) feedbacks is to reduce the voltage gain-bandwidth product. The concept that the gain-bandwidth product is invariant with feedback is therefore only a first-order approximation.

Equations 10, 17, and 25 describe the transient of $\phi_{p(t)}$ arising conceivably not only from changes of the control voltage E_c, but also from changes of any other quantity that appears in these equations. Note that the time constant of the transient is defined by the values of the circuit parameters which exist after the change while the initial value of ϕ_p to be used in equations 10, 17, and 25 is that prevailing before the change.

IV-4. Limitations of Approximate Approach

If finer issues of "half-cyclic lags" and associated phase shifts are important, the approximate treatment described in the foregoing is not adequate. The transition from equation 8 to equation 9 loses the resolution required, and a direct solution of the original difference equations is necessary. The techniques of solving are similar to those of common differential equations of the same order. A disadvantage is seen, at first, because the solution is in the form of a geometric progression rather than an exponential; however, the desirable familiar concept of time constants can be preserved as shown originally by Kikuchi[3] for the basic amplifier. A comprehensive treatment is presented in the following. Again attention is focused on ϕ_p(i.e., on $\phi_{p(n)}$) as the key variable.

IV-5. Modular Differential Equation of Basic Amplifier

(a). *Response to a Step of D-C Control Voltage:* To simplify the writing, it is understood in the following that all variables are increments from some initial steady-state value. Hence the flux level $\phi_{p(n)}$ established at the end of the nth half-cycle is related to the average mmf, $F_{p(n)}$, that acted upon the resetting core during that half-cycle simply by $\phi_{p(n)}=K_\phi F_{p(n)}$, assuming K_ϕ is constant over the range of interest.

In the basic amplifier without bias circuits $F_{p(n)}=N_cI_{c(n)}$, and thus equation 8 becomes

$$E_{c(n)}=\frac{R_c}{K_\phi N_c}\,\phi_{p(n)}+2fN_c(\phi_{p(n)}-\phi_{p(n-1)}) \tag{26}$$

This difference equation is linear with constant coefficients over the range. For a step of d-c signal voltage E_c, at the beginning of the first half-cycle, equation 26 can be rewritten as

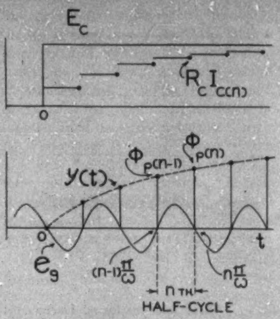


Fig. 3. Transient of preset flux $\phi_{p(n)}$ and control current $I_{c(n)}$

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=\zeta E_c \tag{27}$$

where $\zeta=\dfrac{K\phi\dfrac{N_c}{R_c}}{1+K_{EN}}$ and

$\alpha=K_{EN}{}^*/1+K_{EN}{}^*<1$ and its solution is

$$\phi_{p(n)}=\zeta\,\frac{E_c(1-\alpha^n)}{1-\alpha}=E_cK_\phi N_c/R_c(1-\alpha^n) \tag{28}$$

Fig. 3 shows the transient processes as described by the solution of the difference equation. The concept of a time constant is preserved by writing an auxiliary linear differential equation

$$\frac{dy}{dt}+ay=b \tag{29}$$

and prescribing that its solution will satisfy the condition

$$y(t)=\phi_{p(n)} \text{ at } t=\frac{n\pi}{\omega},\ n=0,\ 1,\ 2,\ 3\ \ldots$$

(In other words, the solution of the modular differential equation 29 is a continuous function, shown by the line which passes through the points expressing the discrete values of ϕ_p in Fig. 3.) The condition is met if

$$a=\frac{\omega}{\pi}\ \ln\frac{1}{\alpha};\quad b=\left(\frac{\omega}{\pi}\ \ln\frac{1}{\alpha}\right)\frac{\zeta E_c}{1-\alpha} \tag{30}$$

The function, $y(t)$ is characterized by a time constant

$$\tau_y=\frac{1}{a}=\frac{\pi}{\omega}\ \frac{1}{\ln 1/\alpha} \tag{31}$$

where $0<\alpha<1$

and the usefulness of this modular time constant to express amplifier transient is evident from Fig. 3.

Since $\alpha=K_{EN}{}^*/(1+K_{EN}{}^*)$, it follows that

$$\ln\frac{1}{\alpha}=\ln\left(\frac{1+K_{EN}{}^*}{K_{EN}{}^*}\right)=2\left[\frac{1}{(2K_{EN}{}^*+1)}+\frac{1}{3}\left(\frac{1}{2K_{EN}{}^*+1}\right)^3+\frac{1}{5}\left(\frac{1}{2K_{EN}{}^*+1}\right)^5\cdots\right]$$

$$\approx 2\left(\frac{1}{2K_{EN}{}^*+1}\right)$$

with an error of less than 4% if $K_{EN}{}^*\geq 1$. Hence

$$\tau_y\approx\frac{\pi}{\omega}\left(K_{EN}{}^*+\frac{1}{2}\right)\text{ seconds} \tag{32}$$

This expression is now compared with the approximate time constant of equation 11, and it is seen that

$$\tau_y\approx\tau_a+\frac{1}{2}\,\frac{\pi}{\omega}\text{ (seconds)} \tag{33}$$

The readjustments of ϕ_p in this approach are slower than are predicted by the approximate time constant τ_a; the two approaches are essentially equivalent if $K_{EN}{}^*$ is large, as should be expected, because then the transient is slow.[3]

In addition, a half-cyclic transportation lag remains inherent to the half-cyclic average output current of the basic amplifier because $I_{L(n)}$ is determined solely by $\phi_{p(n-1)}$. Actually, in questions of fast response, the waveform of the instantaneous i_L may be of significant interest; the angle of firing, however, is easily obtained from the half-cyclic average values.

If a bias circuit is coupled, it can be shown that τ_y is still given by equation 33. In this case, however, τ_a is given by equation 12.

(b). *Sinusoidal Response:* Consider a signal, $e_c=E_{c(\max)}\sin\Omega t$, applied to the original bridge circuit (suitably biased from a d-c source) at the beginning of a half-cycle of the gate voltage. The angular frequency Ω of the signal is smaller than the carrier frequency ω. The values of ϕ_p, as increments from the value due to the bias, are obtained from the difference equation

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=\frac{E_{c(\max)}\zeta\omega}{\pi}\int_{t=(n-1)\pi/\omega}^{t=n\pi/\omega}\sin\Omega t\,dt$$

or

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=\frac{E_{c(\max)}\zeta}{\lambda}[\cos(n-1)\lambda-\cos n\lambda] \tag{34}$$

where $\lambda=(\pi\Omega/\omega)$. Rewriting for simplicity yields

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=C_1[\cos(n-1)\lambda-\cos n\lambda] \qquad (35)$$

where $C_1=E_{c(\max)}\zeta/\lambda$. The solution is

$$\phi_{p(n)}=C_1\left[\frac{\cos\lambda(n-1)-(\alpha+1)\cos\lambda n}{\alpha^2+1-2\alpha\cos\lambda}+\frac{\alpha\cos\lambda(n+1)}{\alpha^2+1-2\alpha\cos\lambda}\right]+\frac{C_1\alpha^n(\alpha+1)(1-\cos\lambda)}{\alpha^2+1-2\alpha\cos\lambda} \qquad (36)$$

The first term is the a-c steady-state solution while the second term represents the transient. Again, a modular differential equation may be written

$$\frac{dy}{dt}+ay=c\sin(\Omega t-\gamma) \qquad (37)$$

The complete solution of equation 37 is

$$y(t)=\frac{c\sin(\Omega t-\gamma-\theta)}{\sqrt{a^2+\Omega^2}}+\left[\frac{c\sin(\theta+\gamma)}{\sqrt{a^2+\Omega^2}}\right]\epsilon^{-at} \qquad (38)$$

where $\theta=\tan^{-1}\Omega/a$

Again it is prescribed $y(t)=\phi_{p(n)}$ for any $t=n\pi/\omega$. Hence, substituting $n\pi/\omega$ for t in equation 38, and equating the steady-state and transient terms to the corresponding terms of equation *36* results in the following relations:

$$a=\frac{\omega}{\pi}\ln\frac{1}{\alpha} \qquad (39)$$

$$c=\frac{C_1[(a+\Omega)(1-\alpha)\sin\lambda+a(1+\alpha)(\cos\lambda-1)]}{\alpha^2+1-2\alpha\cos\lambda} \qquad (40)$$

$$\sin\gamma=\frac{\Omega(\alpha-1)\sin\lambda+a(1+\alpha)(1-\cos\lambda)}{(a+\Omega)(1-\alpha)\sin\lambda+(a-\Omega)(1+\alpha)(\cos\lambda-1)} \qquad (41)$$

In the steady state, $y(t)$ lags behind the signal e_c by an angle $(\gamma+\theta)$ radians of the signal frequency. Having obtained $\phi_{p(n)}$, the half-cyclic average load current of the nth half-cycle is given as usual, by

$$I_{L(n)}=\frac{2fN_g}{R_g}\phi_{p(n-1)} \qquad (42)$$

No direct phase comparison can be established between the continuous signal e_c and the discrete values of half-cyclic average output. The instantaneous values of i_L can be obtained easily from the average half-cyclic values and a phase comparison between e_c and the fundamental (at frequency Ω) of i_L can be made as shown in Fig. 4. It is noted that, with small sinusoidal signals applied to an amplifier biased close to full output, the instantaneous output current has a fundamental component at frequency Ω that lags behind the applied signal e_c by little more than $(\gamma+\theta)$. The lag will approach $(\gamma+\theta+\lambda)$ if the amplifier is biased in the region of low outputs, i.e., a difference of phase shifts by about one half-cycle of the carrier frequency exists in the two cases.

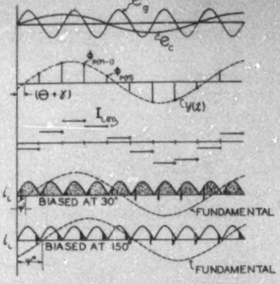


Fig. 4. Phase relationships of fundamental of output versus input for sinusoidal input voltage with two different values of bias

IV-6. EXTENSIONS OF MODULAR APPROACH TO AMPLIFIERS WITH FEEDBACK

Extensions of this approach to feedback amplifiers are possible with the use of the pertinent difference equations of the various loops. Two simple examples of magnetic and electric feedback are treated to illustrate the procedure. The results of all cases are given in Appendixes I through III.

For the amplifier with magnetic load-current feedback, Appendix II(A), equations 8 and 15 may be written. If the gate voltage and the control voltage are constant, considering incremental quantities only, equation 15 reduces to

$$I_{f(n)}=\frac{2fN_gr_s}{R_g'R_f}\phi_{p(n-1)}\pm\frac{2fN_fR_g}{R_g'R_f}(\phi_{p(n)}-\phi_{p(n-1)}) \qquad (43)$$

From the core-control curve, with constant K_ϕ

$$N_cI_{c(n)}\mp N_fI_{f(n)}=\frac{\phi_{p(n)}}{K_\phi} \qquad (44)$$

Substituting equation 43 into equation 44, solving for $I_{c(n)}$, and substituting the result into equation 8 yields

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=\zeta E_c \qquad (45)$$

where

$$\alpha=\frac{K_{BN}^*\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}\frac{R_g}{R_g'}\right)\mp\beta K_B^*}{1+K_{BN}^*\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}\frac{R_g}{R_g'}\right)} \qquad (46)$$

$$\zeta=\frac{K_\phi N_c/R_c}{1+K_{BN}^*\left(1+\frac{N_f^2}{N_c^2}\frac{R_c}{R_f}\frac{R_g}{R_g'}\right)} \qquad (47)$$

and all other symbols are defined in Appendix II(A).

The solution of equation 45 is of the same form as equation 28, i.e.,

$$\phi_{p(n)}=\frac{\zeta E_c(1-\alpha^n)}{1-\alpha} \qquad (48)$$

The load current is obtained by writing the voltage equation of the feedback loop as

$$I_{L(n)}=\left(\frac{R_f+r_s}{r_s}\right)I_{f(n)}\mp\frac{2fN_f}{r_s}\times(\phi_{p(n)}-\phi_{p(n-1)}) \qquad (49)$$

Substituting equation 43 into equation 49 and collecting terms yields

$$I_{L(n)}=C_2\phi_{p(n)}+C_3\phi_{p(n-1)} \qquad (50)$$

where

$$C_2=\pm\frac{2fN_fr_s}{R_fR_g'} \qquad (51)$$

$$C_3=\frac{2_fN_fr_s}{R_fR_g'}\left[\mp1+\frac{N_g}{N_f}\left(1+\frac{R_f}{r_s}\right)\right] \qquad (52)$$

Equation 50 shows that a step change of signal causes a change of load current already in the first half-cycle due to the transient voltage induced in the feedback loop. The modular equation is the same as equation 29 and for $y(t)$ to be equal to $\phi_{p(n)}$ at all $t=n\pi/\omega$, the conditions of equation 30 must be met. For sinusoidal signals, the difference equation of the control circuit is

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=C_1(\cos(n-1)\lambda-\cos n\lambda) \qquad (53)$$

where $C_1=\zeta E_{c(\max)}/\lambda$; λ and $E_{c(\max)}$ have been defined in section IV-5(b). The solution has the same form as equation 36. The modular equation, then, follows the same pattern as the basic, with the proper substitution of the constants pertaining to this type of feedback.

For series electric load-current feedback, Appendix III(A), equations 21 and 22 are written. On an incremental basis for constant gate voltage and constant control voltage, equation 22 becomes

$$I_{L(n)}=\frac{2_fN_g\phi_{p(n-1)}}{R_g}\mp\frac{r_s}{R_g}I_{c(n)} \qquad (54)$$

From the core-control curve, with constant K_ϕ

$$I_{c(n)}=\frac{\phi_{p(n)}}{K_\phi N_c} \qquad (54A)$$

Substituting this relationship and equation 54 into equation 21 yields

$$\phi_{p(n)}-\alpha\phi_{p(n-1)}=\zeta E_c \qquad (55)$$

where

$$\alpha=\frac{K_{EN}^{*}\mp\beta K_{E}^{*}}{1+\nu+K_{EN}^{*}} \tag{56}$$

and

$$\zeta=\frac{K_{\phi}N_c/R_c}{1+\nu+K_{EN}^{*}} \tag{57}$$

The symbols ν and β are defined in Appendix III(A).

The solution of equation 55 is

$$\phi_{p(n)}=\zeta E_c\frac{(1-\alpha^{n})}{1-\alpha} \tag{58}$$

Substituting equation 54(A) into equation 54 gives

$$I_{L(n)}=C_2\phi_{p(n)}+C_3\phi_{p(n-1)} \tag{59}$$

where

$$C_2=\mp\frac{r_s}{R_gK_{\phi}N_c} \tag{60}$$

and

$$C_3=\frac{2fN_g}{R_g} \tag{61}$$

The procedure for the modular equation follows the same pattern as for magnetic feedback.

V. Interpretation of Analytical Results

Appendixes I through III condense the results of a careful analysis which has avoided premature simplifications in the writing of circuit equations and in the process of solving. This has been done intentionally because the influence and implications of the various parameters of the system are not immediately clear in questions of magnetic amplifier feedbacks. A comprehensive formulation of all parameters is needed also as a starting point for future studies of the drift problem. As a consequence, the expressions given are moderately complex. In any specific problem, the numerical values of the various terms may suggest reasonable simplifications, as the case may be.

Some comments on the meaning of the results may be appropriate at this point. (Negative feedbacks are discussed; for positive feedbacks, of course, the trends reverse.)

The output resistance R_o, defined in Fig. 2, decreases for all types of negative load-voltage feedback (magnetic or electric) and increases for all types of negative load-current feedback. (This is the essential difference in the effects of load-voltage and load-current feedback.)

The input resistance, R_i, is not affected by magnetic feedbacks. Electric feedbacks do not modify the control winding current I_c, needed for a given output. But for negative-series electric feedback, the input resistance is increased since a larger control voltage E_c is needed to cause this I_c to flow. A negative shunt electric feedback, on the other hand, causes a larger input current I_i to be supplied from the signal source for a given I_c. The voltage E_c must be increased then to supply the additional drop across R_K; the over-all result is a decrease of R_i from the original value R_K+2r_c to a value $R_K+m\times 2r_c$, where $0<m<1$. (This is a unique feature of this type of feedback.)

The gross voltage gain K_E decreases for any type of negative feedback. The current gain, K_A decreases for negative magnetic feedback at about the same rate as K_E. For negative series electric feedback, K_A remains essentially constant, while for negative shunt electric feedback it decreases, but at a slower rate than K_E. The gross volt-ampere gain $K_{EA}=K_E\times K_A$ therefore decreases in all cases, although at different rates for the different feedbacks.

The approximate time constant τ_a, decreases proportionally to K_E if the delaying action of coupled circuits of bias or feedback is neglected. (In the case of shunt electric feedback, a delaying action arises from the resistive network connected across the control winding terminals.) On this basis, the gross voltage gain-time constant ratio (gain-bandwidth product) is invariant with any amount or type of feedback.

The modular time constant τ_y also decreases with decreasing K_E; for sufficient amounts of negative feedback it may be made smaller than the corresponding τ_a. For instance, when the voltage gain per turn K_{EN} is 1.25, the value of α, neglecting loading and coupling terms, is $\alpha=(1+1/5K_{EN}^{*})/(1+K_{EN}^{*})$. If the basic voltage gain per turn, K_{EN}^{*}, was 20 to start with, $\alpha=5/21$ and thus the modulator time constant τ_y is 7/10 of one half-cycle, while the approximate approach would suggest $\tau_a\approx 1.25$ half-cycles. Likewise for $K_{EN}=1.0$, $\alpha=1/(1+K_{EN}^{*})$, i.e., for $K_{EN}^{*}=20$, $\tau_y\approx 1/3$ of one half-cycle while τ_a is one half-cycle. (At this value of K_{EN}, however, misbehavior of rectifiers may begin to intervene significantly to reduce the accuracy of the analysis as will be seen in the next section.)

The constant C_1 of the Appendixes is used in the difference-equation analysis of sinusoidal response; C_2 and C_3 appear in the general expression for the load current $I_{L(n)}=C_2\phi_{p(n)}+C_3(\phi_{p(n-1)})$.

For quick reference, the qualitative effects of the various feedback are summed up in Appendix IV. As a final remark, it may be noted that the introduction of a forcing resistor in series with the control signal of the basic amplifier has effects similar to those of a negative series electric feedback in terms of gross voltage gain, time constant τ_a, and dynamic gain, D_p. But, of course, the series electric feedback has other features: it changes the output resistance R_o; it can stabilize an amplifier made unstable by other forms of feedback; it is helpful in linearizing the output-voltage–input-voltage characteristics over ranges of output for which K_{ϕ} is not constant and can make τ_y smaller than τ_a as seen in the foregoing.

VI. Limitations of the Analysis

Experimentation on a number of amplifiers with various feedbacks has been performed to establish the limits of validity of the assumptions.

Measured transfer characteristics of gross output voltage versus average reset mmf and also detailed observation of the wave form of instantaneous core reset voltages and reset mmf's proved that the assumption of a "core-control" curve was actually a basis quite tenable and adequate for the analytical predictions. The application of negative feedbacks, electric or magnetic, slows down the rate of reset in the prefiring interval, and increases it in the postfiring interval. These two effects tend to compensate each other fairly well in terms of the average value of the mmf of reset. Experimental transfer characteristics and oscillograms may be found in reference 2; they are omitted here. (It may be noted parenthetically that the slopes of the core-control curve depend on the gate voltage roughly proportionally over the range ±20% of normal.)

Departures from the analytical predictions began to occur when the negative feedback was increased beyond certain limits. A large negative feedback requires that the d-c signal voltage E_c be positive and large, especially at high outputs. Under these conditions, the rectifier associated with the gating core may block during an appreciable part of the prefiring interval. During this time, the resetting core remains at positive saturation, and the instantaneous control current i_c is positive, being dictated by the mmf requirements of the ascending core. All this results in a somewhat more positive value of the average mmf of reset F_p for a given output, i.e., the transfer characteristic of gross output voltage E versus F_p exhibits a pronounced droop at higher outputs.

In a typical bridge amplifier with Hypernik V, toroidal tape cores, $1^{1}/_{2}$

by 2 inches by 1/2 inch, 2 mils, and with $N_c=400$, $N_g=3{,}000$, $R_c=170$ ohms, $r_{gr}=130$ ohms, $R_g=1{,}400$ ohms at 60 cycles per second and normal gate voltage, this droop became noticeable for amounts of negative series electric or magnetic feedback which made the gross voltage gain per turn, $K_{EN}=1.25$. In terms of a transfer characteristic expressing output voltage versus input voltage, the effect is hardly noticeable because of the linearizing action of the negative feedback. The rectifier blocking effect is mentioned as a first type of departure from the assumptions of the analysis. (This effect does not occur for a-c rectified signal voltages in phase with the gate voltage as long as $K_{EN}\geq 1$.)

If the negative feedback is increased further, a more serious effect arises. The rate of reset of the resetting core $N_c\dot{\phi}$ tries to exceed the gate-limiter voltage ($-e_g$ for the bridge, and $-(1+R_L/R_g)e_g$ for the center tap), and the rectifier associated with the resetting core unblocks. Then a positive current flows in the gate winding of the resetting core which requires a compensating negative addition to the control current. These effects appear first in the region of low outputs which are associated with comparatively large swings of fluxes. This is a type of spurious positive feedback mentioned in section III-4(e), which distorts significantly the linearity of the output-input voltage characteristic. Practically, these effects occur when the voltage gain per turn becomes less than one.

These two effects, blocking and unblocking of rectifiers at improper times, violate the assumptions of the analysis. It is seen that their appearance is associated with rather extreme amounts of feedback. Experimentation on transient processes by the methods of reference 4 and by observations of rectifier voltages, failed to show any cases of rectifier misbehaviors if they did not occur in the steady state over the same range. The agreement between predicted and observed values of the modular time constant extended to values of τ_y as short as 1/4 cycle.

These considerations apply for bridge and center-tap circuitries. On the other hand, the doubler circuitries appeared particularly vulnerable to rectifier unblocking. For example, in the amplifier mentioned, in doubler connection, an amount of negative feedback, electric or magnetic, that reduced the voltage per turn from the original value, 15, to about half of that value, caused pronounced unblocking effects. This vulnerability of the doubler is due to the low value of the gate-limiter voltage as is generally known.

III. Conclusions

An analysis of 2-core full-wave self-saturating magnetic amplifiers modified by various types of electric and magnetic feedbacks has been presented. Expressions for steady-state gains and input and output resistances have been derived in terms of the gross voltage gain of the bridge amplifier, taken as the basis of reference.

With the preset flux level ϕ_p as the key variable, the transient analysis has been performed. Two different methods are employed. First, the familiar approach in which the half-cyclic average quantities are considered as continuously varying time functions has been extended to the various types of feedback circuitries. The results are given in the form of an approximate time constant τ_a. These expressions show the coupling effects inherent to magnetic and shunt electric feedbacks which are similar to that of a bias circuit coupling. The concept of the voltage gain-bandwidth product being invariant with any type and amount of feedback is seen to be only a first-order approximation.

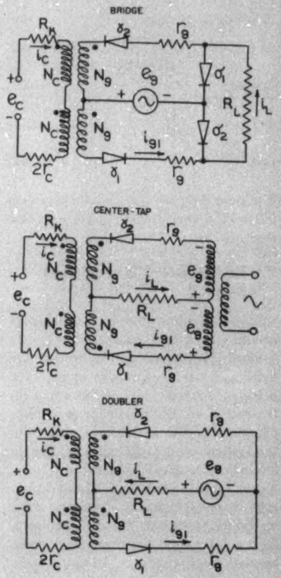
When questions of fast response arise, this approach is inadequate, and a direct solution of the difference equations is necessary. The results thus obtained include findings derived by Hubbard[5] with block diagram techniques using similar physical assumptions. The solution of the difference equation leads to a geometric progression for $\phi_{p(n)}$ but the use of an auxiliary exponential time function $y(t)$ preserves the concept of a simple time constant to describe the values of ϕ_p at the end of each half-cycle of a transient. This modular time constant τ_y has been given for the various feedbacks considered.

The half-cyclic load current $I_{L(n)}$ is obtained from the values of preset flux levels $\phi_{p(n-1)}$ and $\phi_{p(n)}$. For the basic amplifier, the half-cyclic average load current $I_{L(n)}$ depends solely on $\phi_{p(n-1)}$. Modifications by feedback introduce an additional dependency on $\phi_{p(n)}$, i.e., already in the first half-cycle the average load current is changed by some amount. In most practical cases, however, this is a minor effect and the concept of a half-cyclic time lag remains tenable for the load current.

The experiment shows that departures from the analytical predictions occurred when rectifiers blocked or unblocked at improper times in the half-cycle. In the bridge and center tap, these effects do not appear significantly, either in the steady state or transient, as long as $\tau_y \geq 1/4$ cycle. On the other hand, in the doubler, rectifier unblocking occurs for very small amounts of negative feedback, electric or magnetic. The analysis of unblocking effects in the steady state and transient requires more refined magnetic information; work is in progress along these lines for some simple cases.

Appendix I. Basic Amplifiers

The bridge, center-tap, and doubler circuits are shown in the following figure:



$$R_i=R_c$$

$$R_g=R_L+r_{gr}$$

$$R_o=r_{gr}=r_g+r_\gamma+r_\sigma$$

$$K_E{}^*=2fN_gK_\phi\frac{N_c}{R_c}$$

$$\mathcal{K}=K_E{}^*$$

$$\tau_a=\frac{1}{2f}K_E{}^*\frac{N_c}{N_g}=\frac{1}{2f}K_{EN}{}^*$$

$$\alpha=\frac{K_{EN}{}^*}{1+K_{EN}{}^*}$$

$$\tau_y=\frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)}\approx\tau_a+\frac{1}{2f}\frac{1}{2}$$

$$C_1=\frac{\omega E_{c(\max)}K_\phi \dfrac{N_c}{R_c}}{\pi\Omega(1+K_{EN}{}^*)}$$

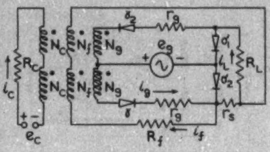
$$C_2=0$$

$$C_3=2f\frac{N_g}{R_g}$$

Appendix II. Magnetic Feedbacks Including Special Types

Magnetic Feedbacks

A—LOAD CURRENT



$$K_E=\frac{K_E{}^*}{1\pm\beta K_E{}^*}$$

$$R_1=R_c$$

$$R_o=r_{gr}+\frac{r_s\left(1\pm\dfrac{R_cN_f}{N_cR_f}K_E{}^*\right)}{1+\dfrac{r_s}{R_f}}$$

$$\mathcal{K}=K_E{}^*$$

where

$$\beta=\frac{R_c}{N_c}\,\frac{N_f}{R_f}\,\frac{r_s}{R_g'}$$

$$R_g'=R_g+(R_L+r_{gr})\frac{r_s}{R_f}$$

$$R_g=R_L+r_{gr}+r_s$$

$$r_{gr}=r_g+r_\gamma+r_\sigma$$

$$\tau_a=\frac{1}{2f}K_{EN}\left(1+\frac{N_f{}^2}{N_c{}^2}\,\frac{R_c}{R_f}\,\frac{R_g}{R_g'}+\frac{N_b{}^2}{N_c{}^2}\,\frac{R_c}{R_b}\right)$$

$$\alpha=\frac{K_{EN}{}^*\left(1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g}{R_g'}+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right)\mp\beta K_E{}^*}{1+K_{EN}{}^*\left(1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g}{R_g'}+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right)}$$

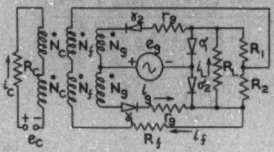
$$\tau_y=\frac{1}{2f}\,\frac{1}{\ln\left(\dfrac{1}{\alpha}\right)}\quad 0\le\alpha<1$$

$$C_1=\frac{\omega E_{c(\max)}K_\phi\dfrac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}{}^*\left(1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g}{R_g'}+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right)\right]}$$

$$C_2=\pm\frac{2fN_fr_s}{R_fR_g'}$$

$$C_3=\frac{2fN_fr_s}{R_fR_g'}\left[\frac{N_g}{N_f}\left(1+\frac{R_f}{r_s}\right)\mp1\right]$$

B—LOAD VOLTAGE



$$K_E=\frac{K_E{}^*}{1\pm\beta K_E{}^*}$$

$$R_1=R_c$$

$$R_o=\cfrac{r_{gr}}{1+\cfrac{r_{gr}}{\left(R_1+\cfrac{R_2R_f}{R_2+R_f}\right)}\pm\cfrac{N_fR_2R_cK_E{}^*}{N_cR_1(R_2+R_f)+R_2R_f}}$$

$$\mathcal{K}=K_E{}^*\frac{R_o}{r_{gr}}$$

where

$$R_g'=R_g(R_1+R_2)+r_{gr}R_L$$

$$R_g''=R_g'+\frac{R_2}{R_f}(R_gR_1+r_{gr}R_L)$$

$$R_g=R_L+r_{gr}$$

$$r_{gr}=r_g+r_\gamma+r_\sigma$$

$$\beta=\frac{N_f}{N_c}\,\frac{R_cR_LR_2}{R_fR_g''}$$

$$\tau_a=\frac{1}{2f}K_{EN}\left(1+\frac{N_f{}^2}{N_c{}^2}\frac{R_c}{R_f}\,\frac{R_g'}{R_g''}+\frac{N_b{}^2}{N_c{}^2}\frac{R_c}{R_b}\right)$$

$$\alpha=\frac{K_{EN}{}^*\left[1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g'}{R_g''}+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right]\mp\beta K_E{}^*}{1+K_{EN}{}^*\left(1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g'}{R_g''}+\dfrac{N_b{}^2}{N_c}\dfrac{R_c}{R_b}\right)}$$

$$\tau_y=\frac{1}{2f}\,\frac{1}{\ln\left(\dfrac{1}{\alpha}\right)}\quad 0\le\alpha<1$$

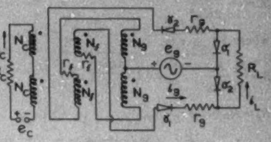
$$C_1=\frac{\omega E_{c(\max)}K_\phi\dfrac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}\left(1+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_f}\dfrac{R_g'}{R_g''}+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right)\right]}$$

$$C_2=\mp\frac{2fN_fr_{gr}R_2}{R_fR_g''}$$

$$C_3=\frac{2fN_fr_{gr}R_2}{R_fR_g''}\times\left(\frac{R_fR_g''(R_1+R_2)N_g}{r_{gr}R_g'R_2N_f}-\frac{N_g}{N_f}\,\frac{R_LR_2}{R_g'}\pm1\right)$$

Special Types of Magnetic Feedback

C—CROSS-CONNECTED FEEDBACK WINDINGS



$$K_E=\frac{K_E{}^*\left(1\pm\dfrac{N_f}{N_g}\right)}{1\pm\dfrac{N_f}{N_c}\dfrac{R_c}{R_g}\left(1\pm\dfrac{N_f}{N_g}\right)K_E{}^*}$$

$$R_1=R_c$$

$$R_o=r_f+r_{gr}+\frac{N_f}{N_c}R_c\left(1\pm\frac{N_f}{N_g}\right)K_E{}^*$$

$$\mathcal{K}=\left(1\pm\frac{N_f}{N_g}\right)K_E{}^*$$

where

$$R_g=R_L+r_{gr}+r_f$$

$$\tau_a=\frac{1}{2f}K_{EN}\left(1\mp\frac{N_f}{N_g\pm N_f}+\frac{N_b{}^2}{N_c{}^2}\frac{R_c}{R_b}\right)$$

$$\alpha=\frac{K_{EN}{}^*\left(1+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}\right)\mp\dfrac{R_c}{R_g}\dfrac{N_f}{N_c}K_E{}^*}{1+K_{EN}{}^*\left(1+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_g}\right)}$$

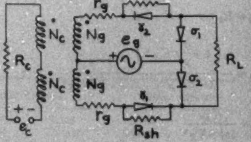
$$\tau_y=\frac{1}{2f}\,\frac{1}{\ln\left(\dfrac{1}{\alpha}\right)}\quad 0\le\alpha<1$$

$$C_1=\frac{\omega E_{c(\max)}K_\phi\dfrac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}{}^*\left(1+\dfrac{N_b{}^2}{N_c{}^2}\dfrac{R_c}{R_b}+\dfrac{N_f{}^2}{N_c{}^2}\dfrac{R_c}{R_g}\right)\right]}$$

$$C_2=\pm2f\frac{N_f}{R_g}$$

$$C_3=2f\frac{N_g}{R_g}$$

D—NEGATIVE FEEDBACK OF SHUNTED GRATE RECTIFIERS (BRIDGE)



$$R_{sh}>>R_L;\ R_L>>r_{gr}$$

$$K_E=\frac{K_E{}^*}{1+\beta\dfrac{R_L}{R_g}K_E{}^*}$$

$$R_i = R_c$$

$$R_o = r_g + r_\gamma + \frac{r_\sigma}{1+\beta K_E^*}$$

$$\mathcal{K} = \frac{K_E^*}{1+\beta K_E^*}$$

where

$$\beta = \frac{N_g}{N_c}\frac{R_c}{R_{sh}}$$

$$R_g = R_L + r_g + r_\sigma + r_\gamma$$

$$\tau_a = \frac{1}{2f} K_{EN}\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)$$

$$\alpha = \frac{K_{EN}^*\left[1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right]-\beta K_E^*}{1+K_{EN}^*\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)}$$

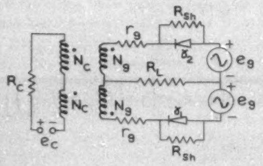
$$\tau_y = \frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)} \qquad 0 \le \alpha < 1$$

$$C_1 = \frac{\omega E_{c(\max)} K_\phi \frac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}^*\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)\right]}$$

$$C_2 = \frac{-2fN_g r_\sigma}{R_g R_{sh}}$$

$$C_3 = 2f\frac{N_g}{R_g}$$

E—NEGATIVE FEEDBACK OF SHUNTED GATE RECTIFIERS (CENTER TAP)



$$R_{sh} >> R_L \qquad R_L >> r_{gr}$$

$$K_E = \frac{K_E^*}{1+\beta\frac{R_L}{R_g}K_E^*}$$

$$R_i = R_c$$

$$R_o = \frac{r_s\left(1+\frac{1}{2}\beta K_E^*\right)}{1+\beta K_E^*} + \frac{r_\gamma}{(1+\beta K_E^*)}$$

$$\mathcal{K} = \frac{K_E^*}{1+\beta K_E^*}$$

where

$$\beta = 2\frac{R_c}{N_c}\frac{N_g}{R_{sh}}$$

$$r_{gr} = r_g + r_\gamma$$

$$R_g = R_L + r_{gr}$$

$$\tau_a = \frac{1}{2f} K_{EN}\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)$$

$$\alpha = \frac{K_{EN}^*\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)-\beta K_E^*}{1+K_{EN}^*\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)}$$

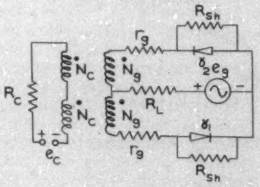
$$\tau_y = \frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)} \qquad 0 \le \alpha < 1$$

$$C_1 = \frac{\omega E_{c(\max)} K_\phi \frac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}^*\left(1+\beta\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)\right]}$$

$$C_2 = \frac{-2fN_g r_{gr}}{2R_L R_{sh}}$$

$$C_3 = \frac{2fN_g}{R_g}$$

F—NEGATIVE FEEDBACK OF SHUNTED GATE RECTIFIERS (DOUBLER)



$$K_E = \frac{K_E^*}{1+\beta K_E^*}$$

$$R_i = R_c$$

$$R_o = \frac{(r_g + R_{sh} + \beta R' K_E^*) r_{gr}}{2r_g + R_{sh} + r_\gamma}$$

$$\mathcal{K} = K_E^*$$

where

$$\beta = \frac{R_c}{N_c}\frac{N_g}{R'}$$

$$r_{gr} = r_g + r_\gamma$$

$$R_g = R_L + r_{gr}$$

$$R' = R_L + \frac{R_g(r_g + R_{sh})}{r_{gr}}$$

$$\tau_a = \frac{1}{2f} K_{EN}\left(1+\beta\frac{R_g}{r_{gr}}\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)$$

$$\alpha = \frac{K_{EN}^*\left(1+\beta\frac{R_g}{r_{gr}}\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)-\beta K_E^*}{1+K_{EN}^*\left(1+\beta\frac{R_g}{r_{gr}}\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)}$$

$$\tau_y = \frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)} \qquad 0 \le \alpha < 1$$

$$C_1 = \frac{\omega E_{c(\max)} K_\phi \frac{N_c}{R_c}}{\pi\Omega\left[1+K_{EN}^*\left(1+\beta\frac{R_g}{r_{gr}}\frac{N_g}{N_c}+\frac{N_b^2}{N_c^2}\frac{R_c}{R_b}\right)\right]}$$

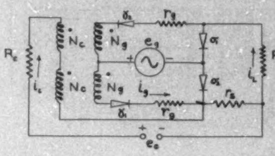
$$C_2 = \frac{2fN_g}{R'}$$

$$C_3 = 2f\frac{N_g}{R'}\frac{(R_{sh}+r_g)}{r_{gr}}$$

Appendix III. Series and Shunt Electric Feedback

Series Electric Feedback

A—LOAD CURRENT



$$K_E = \frac{K_E^*}{1+\nu \pm \beta K_E^*}$$

$$R_i = \frac{K_E^*}{K_E} R_c$$

$$R_o = r_{gr} + \frac{r_s(1 \pm K_E^*)}{1+\frac{r_s}{R_c}}$$

$$\mathcal{K} = \frac{K_E^* - \frac{r_s}{R_c}}{1+\frac{r_s}{R_c}}$$

where

$$\beta = \frac{r_s}{R_g}$$

$$\nu = \frac{r_s\left(1-\frac{r_s}{R_g}\right)}{R_c}$$

$$R_g = R_L + r_s + r_{gr}$$

$$r_{gr} = r_g + r_\gamma + r_\sigma$$

$$\tau_a = \frac{1}{2f} K_{EN}\left(1+\frac{R_c(1+\nu)}{R_b}\frac{N_b^2}{N_c^2}\right)$$

$$\alpha = \frac{K_{EN}^*\left(1+\frac{R_c(1+\nu)N_b^2}{R_b N_c^2}\right) \mp \beta K_E^*}{1+\nu+K_{EN}^*\left(1+\frac{R_c(1+\nu)N_b^2}{R_b N_c^2}\right)}$$

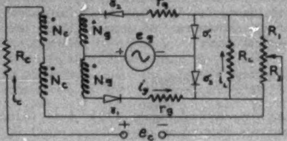
$$\tau_y = \frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)} \qquad 0 \le \alpha < 1$$

$$C_1 = \frac{\omega E_{c(\max)} K_\phi \frac{N_c}{R_c}}{\pi\Omega\left[1+\nu+K_{EN}^*\left(\frac{R_c(1+\nu)N_b^2}{R_b N_c^2}\right)\right]}$$

$$C_2=\mp\frac{r_s}{R_gK_\phi N_c}$$

$$C_3=2f\frac{N_g}{R_g}$$

B—LOAD VOLTAGE



Approximations:
$R_1+R_2>>R_L$ $R_L>>r_{gr}$

$$K_E=\frac{K_E{}^*}{1+\nu\pm\beta\frac{R_L}{R_g}K_E{}^*}$$

$$R_i=\frac{K_E{}^*}{K_E}R_c$$

$$R_o=\frac{r_{gr}(1+\nu)}{1+\nu\pm\beta K_E{}^*}$$

$$K=\frac{K_E{}^*+\beta\frac{r_{gr}}{R_c}}{1+\nu+\beta K_E{}^*}$$

where

$$\beta=\frac{R_2}{R_1+R_2}$$

$$\nu=\beta\frac{R_1}{R_c}$$

$$R_g=R_L+r_{gr}$$

$$r_{gr}=r_g+r_\gamma+r_\sigma$$

$$r_a=\frac{1}{2f}K_{EN}\left(1+\frac{R_c(1+\nu)N_b{}^2}{R_bN_c{}^2}\right)$$

$$\alpha=\frac{K_{EN}{}^*\left[1+\frac{R_c(1+\nu)N_b{}^2}{R_bN_c{}^2}\right]\mp\beta\frac{R_L}{R_g}K_E{}^*}{1+\nu+K_{EN}{}^*\left(1+\frac{R_c(1+\nu)N_b{}^2}{R_bN_c{}^2}\right)}$$

$$r_y=\frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)}\qquad 0\leq\alpha<1$$

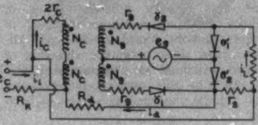
$$C_1=\frac{\omega E_{c(\max)}K_\phi\frac{N_c}{R_c}}{\pi\Omega\left[1+\nu+K_{EN}{}^*\left(1+\frac{R_c(1+\nu)N_b{}^2}{R_bN_c{}^2}\right)\right]}$$

$$C_2=\pm\frac{\beta r_{gr}K_\phi N_c}{R_g+\frac{r_{gr}R_L}{R_1+R_2}}$$

$$C_3=\frac{2fN_g}{R_g+\frac{r_{gr}R_L}{R_1+R_2}}$$

Shunt Electric Feedback

C—LOAD CURRENT



$R_L>>r_s$ $R_L>>r_{gr}$

$$K_E=\frac{K_E{}^*\left(\frac{R_E-R_K}{R_E}\right)}{1-\nu\pm\beta K_E{}^*}$$

$$R_i=R_K+\frac{2r_c(R_E-R_K)}{R_E-R_K+2r_c\pm R_c\frac{r_s}{R_g}K_E{}^*}$$

$$R_o=r_{gr}+r_s\left(1\pm\frac{R_KK_E{}^*}{R_E(1-\nu)}\right)$$

$$\mathcal{K}=K_E{}^*\frac{\left(1-\frac{R_K}{R_E}\right)}{1-\nu}$$

where

$$\beta=\frac{R_K}{R_E}\frac{r_s}{R_g}$$

$$\nu=\frac{R_K{}^2}{R_cE_E}$$

$$R_c=R_K+2r_c$$

$$R_g=R_L+r_{gr}+r_s$$

$$r_{gr}=r_g+r_\gamma+r_\sigma$$

$$R_E=R_a+R_K+r_s$$

$$r_a=\frac{1}{2f}K_{EN}\left(1+\frac{R_K}{R_E-R_K}\right)\times\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{R_c(1-\nu)}{R_b}\right)$$

$$\alpha=\frac{K_{EN}{}^*\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{R_c(1-\nu)}{R_b}\right)\mp\beta K_E{}^*}{1-\nu+K_{EN}{}^*\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{R_c(1-\nu)}{R_b}\right)}$$

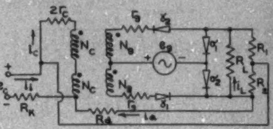
$$r_y=\frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)}\qquad 0\leq\alpha<1$$

$$C_1=\frac{\omega E_{c(\max)}\frac{R_E-R_K}{R_E}K_\phi\frac{N_c}{R_c}}{\pi\Omega\left[1-\nu+K_{EN}{}^*\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{R_c(1-\nu)}{R_b}\right)\right]}$$

$$C_2=\frac{\pm\left[1-\frac{R_K}{R_c}+K_{EN}{}^*\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{2r_c}{R_b}\right)\right]}{\left(R_g+\frac{R_aR_g}{r_s}\right)K_\phi\frac{N_c}{R_c}}$$

$$C_3=\frac{K_E{}^*\left(\frac{R_a}{r_s}+1\right)\mp K_{EN}{}^*\left(1+\frac{N_b{}^2}{N_c{}^2}\frac{2r_c}{R_b}\right)}{\left(R_g+\frac{R_aR_g}{r_s}\right)K_\phi\frac{N_c}{R_c}}$$

D—LOAD VOLTAGE



$R_1+R_2>>R_L$ $R_L>>r_{gr}$

$$K_E=\frac{K_E{}^*\left(\frac{R_E-R_K}{R_E}\right)}{1-\nu\pm\beta\frac{R_L}{R_g}K_E{}^*}$$

$$R_i=R_K+\frac{2r_c(R_E-R_K)}{R_E-R_K+2r_c\pm\frac{R_cR_2}{R_1+R_2}K_E{}^*}$$

$$R_o=\frac{r_{gr}(1-\nu)}{1-\nu\pm\beta K_E{}^*}$$

$$\mathcal{K}=\frac{K_E{}^*\left(\frac{R_E-R_K}{R_E}\right)}{1-\nu\pm\beta K_E{}^*}$$

where

$$\beta=\left(\frac{R_2}{R_1+R_2}\right)\frac{R_K}{R_E}$$

$$\nu=\frac{R_K{}^2}{R_cR_E}$$

$$R_c=R_K+2r_c$$

$$R_g=R_L+r_{gr}$$

$$r_{gr}=r_g+r_\gamma+r_\sigma$$

Table I

	Magnetic		Series Electric		Shunt Electric	
	Load Current	Load Voltage	Load Current	Load Voltage	Load Current	Load Voltage
Output resistance R_o	increase	decrease	increase	decrease	increase	decrease
Input resistance R_i	no change	no change	increase	increase	decrease	decrease
Gross voltage gain K_E	decrease	decrease	decrease	decrease	decrease	decrease
Current gain K_A	decrease with K_E	decrease with K_E	no change	no change	decrease	

$$R_E = R_a + R_K + R_2$$

$$\tau_a = \frac{1}{2f} K_{EN}\left[1+\frac{R_K}{R_E-R_K}\right]\times \left[1+\frac{N_b^2}{N_c^2}\frac{R_c(1-\nu)}{R_b}\right]$$

$$\alpha = \frac{K_{EN}^*\left[1+\frac{N_b^2}{N_c^2}\frac{R_c(1-\nu)}{R_b}\right]\mp\beta\frac{R_L}{R_g}K_E^*}{1-\nu+K_{EN}^*\left[1+\frac{N_b^2}{N_c^2}\frac{R_c(1-\nu)}{R_b}\right]}$$

$$\tau_y = \frac{1}{2f}\frac{1}{\ln\left(\frac{1}{\alpha}\right)} \qquad 0\le\alpha<1$$

$$C_1 = \frac{\omega E_{c(\max)}\frac{R_E-R_K}{R_E}K_\phi\frac{N_c}{R_c}}{\pi\Omega\left[1-\nu+K_{EN}^*\left(1+\frac{N_b^2}{N_c^2}\frac{R_c(1-\nu)}{R_b}\right)\right]}$$

$$C_2 = \frac{\mp\left[1-\frac{R_K}{R_c}+K_{EN}^*\left(1+\frac{N_b^2}{N_c^2}\frac{2r_c}{R_b}\right)\right]}{\left[\frac{R_aR_g(R_1+R_2)}{r_{gr}R_2}+\frac{R_2R_L}{R_1+R_2}\right]K_\phi\frac{N_c}{R_c}}$$

$$C_3 = \frac{K_E^*\frac{R_a}{r_{gr}}\left(\frac{R_1+R_2}{R_2}\right)\pm K_{EN}^*\left(1+\frac{N_b^2}{N_c^2}\frac{2r_c}{R_b}\right)}{\left[\frac{R_aR_g}{r_{gr}}\left(\frac{R_1+R_2}{R_2}\right)+\frac{R_2R_L}{R_1+R_2}\right]K_\phi\frac{N_c}{R_c}}$$

Appendix IV. Primary Effects of Negative Feedback

Table I shows the primary effects of negative feedbacks.

Neglecting delaying effects of coupled bias and feedback loops, the approximate time constant is

$$\tau_a = \frac{1}{2f}K_E\frac{N_c}{N_g} \text{ seconds}$$

The dynamic power gain (figure of merit) is

$$D_p = \frac{K_EK_A}{\tau_a} = 2f\frac{N_g}{N_c}K_A$$

References

1. On Feedback in Magnetic Amplifiers. Part II—Combined Magnetic and Electrical Feedbacks, **L. A. Finzi, J. J. Suozzi.** *AIEE Transactions*, vol. 78, pt. I, 1959 (*Paper no. 58-1232*).

2. On Feedback in Magnetic Amplifiers, **J. J. Suozzi.** *Doctoral dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., July 1958.

3. Analysis of the Full-Wave Magnetic Amplifier Circuits, **Tadashi Kikuchi.** *AIEE Transactions*, vol. 76, pt. I, July 1957, pp. 241–49.

4. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, **D. L. Critchlow.** *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 607–10.

5. Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques, **Robert N. Hubbard.** *Ibid.*, vol. 76, pt. I, Nov. 1957, pp. 578–88.

Discussion

Ernest N. Costogue (Stavid Engineering Company, Plainfield, N. J.): The authors define and classify different types of feedback including resistance-shunting of rectifiers, and the common load-type feedbacks. How does resistance shunting of separate windings on magnetic-amplifier cores fit into this over-all picture, and would this be classified as a type of feedback?

R. O. Decker (Westinghouse Electric Corporation, Cheswick, Pa.): The authors are to be congratulated for their clear treatment of amplifier phenomena under transient conditions. The use of a modular time constant τ_y has merit in conveniently describing discrete values of the preset flux level ϕ_p which the authors take as the key variable of amplifier operation during transient.

The expression

$$\tau_y = \frac{\pi}{\omega}\frac{1}{\ln 1/\alpha}$$

specifies that the auxiliary variable α should be limited to values between zero and unity. When $\alpha=0$, τ_y becomes zero, and a step signal applied at the beginning of a gate supply half-cycle results in the new steady-state value of ϕ_p being established at the end of that same half-cycle. Evidently this is as fast as the amplifier can respond. It would seem, however, that negative values of α may also be obtained in practice if a sufficiently high negative feedback is introduced to modify the basic amplifier. What does this mean in terms of transient readjustments of ϕ_p and of the associated amplifier response?

L. A. Finzi and **J. J. Suozzi**: Mr. Costogue asks whether the resistance shunting of two separate windings on the two cores should be considered as a feedback and inquires about its classification within the general terminology of section I. Assume that the basic amplifier of Appendix I has been modified by adding a coil with N_a turns on each core; each coil is short-circuited directly by a resistor R_a.

Under these conditions the half-cyclic average mmf of the resetting core is modified by a term, N_aI_a. With the convention that the instantaneous current i_a is positive when flowing "into the dot," the half-cycle average mmf of the resetting core is $F_p=N_cI_c+N_aI_a$, where

$$I_a = \frac{2fN_a(\phi_s-\phi_p)}{R_a} = (E_g-E)\frac{N_a}{N_g}\frac{1}{R_a}$$

It may be said that the output variable performing feedback is the gross output voltage E, and that this feedback is negative, since a larger positive increment ΔE_c is needed now to obtain a certain increase ΔE of gross output voltage. Finally, this feedback action is obtained through windings other than the signal winding. Concluding, with the terminology of section I, this is a "negative magnetic gross-output-voltage feedback." The gain K_E is given by

$$K_E = \frac{K_E^*}{1+\beta K_E^*}$$

where

$$\beta = \frac{N_c}{N_g}\frac{R_c}{N_c^2}\Big/\frac{R_a}{N_a^2}$$

This type of feedback does not modify either the input resistance R_i or the output resistance R_o. With a simple extension of the treatment of section IV-2, the familiar result

$$\tau_a = \frac{1}{2f}K_E\frac{N_c}{N_g} \text{ seconds}$$

can be derived again. The fact that currents induced in closed loops result in a faster response may seem strange. However, the present situation should not be confused with that of a bias circuit with low resistance. In the latter case the two coils are connected in series and the induced currents have zero half-cyclic value in the steady state; the bias circuit does not affect the steady-state gain and makes itself obnoxious during the transient. In the case at hand, "induced currents" of a finite half-cyclic value flow in the steady state, and the associated reduction of voltage gain requires larger values of the swing of a signal voltage, for a wanted swing of steady-state output, to start with.

Mr. Decker notices that negative values of α are conceivable, if the amount of negative feedback is sufficiently high. The analysis is easily extended to such unusual situations, because the general difference equation

$$\phi_{p(n)} - \alpha\phi_{p(n-1)} = \zeta E_c$$

and its solution

$$\phi_{p(n)} = \frac{\zeta E_c}{1-\alpha}(1-\alpha^n) = \phi_{p\infty}(1-\alpha^n)$$

remains valid also for negative values of α. In this case, with $\alpha=-1\times|\alpha|$, the solution is rewritten in the form

$$\phi_{p(n)} = \phi_{p(\infty)}[1-(-1)^n|\alpha|^n] = \phi_{p(\infty)}[1\mp|\alpha|^n]$$

where the minus sign applies for $n=2, 4, 6$, etc. and the plus sign applies for $n=1, 3, 5$, etc. As seen in Fig. 5, $\phi_{p(2)}$, $\phi_{p(4)}$, etc., are obtained directly from the modular function, $y(t)$; its time constant may be written as

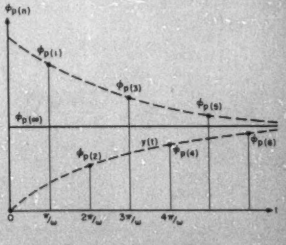


Fig. 5. Transient of preset fluxes $\phi_{p(n)}$ for $\alpha<0$

Vol
Pt.
19

$$\tau_y' = \frac{\pi}{\omega} \frac{1}{\ln 1/|\alpha|}$$

On the other hand, $\phi_{p(1)}$, $\phi_{p(2)}$, etc., are located on a curve which is symmetrical to $y(t)$ about the horizontal line, $\phi_p = \phi_{p(\infty)}$.

This behavior is explained by the general recognition that during the first half-cycle, the output remains essentially the original one and no feedback action intervenes to counterbalance the change of signal. In the limiting case $\alpha = 0$, the signal drives the resetting core exactly to $\phi_{p(\infty)}$ within the first half-cycle, and from then on steady state prevails, since the negative feedback intervenes in the subsequent half-cycles to prevent further changes of ϕ_p. When even larger amounts of negative feedback are used (so as to make $\alpha < 0$), larger values of E_c are needed for the amplifier control, and in the first half-cycle of the transient, the signal E_c "overdoes it," driving ϕ_p to a value $\phi_{p(1)}$ beyond that of the final steady state. A transient of this sort, from a lower to a higher output, is displayed in the tracing of an oscillogram; see Fig. 6. The cores and windings of the basic 60-cycle bridge amplifier are those described in section VI of the paper. The control circuit resistance is 150 ohms, and the original gains are $K_E^* = 140$ and $K_{EN}^* = 18.6$. Negative series-electric output-current feedback, Appendix III(A), has been applied with $r_s = 400$ ohms and $R_g = 1{,}500$ ohms. Hence, $\beta = 0.266$, $\gamma = 1.95$, and $\alpha \cong -0.865$. Accordingly

$$\tau_y' = \frac{1}{2f} \frac{1}{\ln 1/|\alpha|} = 0.057 \text{ seconds}$$

i.e., somewhat less than 7 half-cycles. The effects discussed in section VI for high values of negative feedback limit somewhat the validity of the analysis, but the agreement between this analytical prediction and the oscillographic results shown is still quite good.

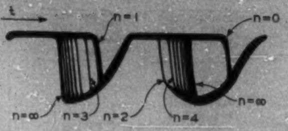


Fig. 6. Wave forms of load voltage in a transient from low to high outputs. Negative series-electric output-current feedback with $\alpha \approx -0.865$

A General Approach for Obtaining Transient Response by the Use of a Digital Computer

P. E. LEGO
ASSOCIATE MEMBER AIEE

T. W. SZE
MEMBER AIEE

OBTAINING the transient response of linear systems is fundamentally a problem involving the solution of the differential equations which describe the behavior of the system. These equations can be solved by the use of classical methods, operational calculus, or by applying the Laplace transformation.

It is not always possible, however, to obtain the equations describing the system response, or if obtainable these equations may be extremely difficult to solve analytically. For example, the inverse Laplace transform requires the determination of the roots of the characteristic equation of the differential equations. For feedback systems with transport delay in the forward loop, difference differential equations are produced which include exponential terms or exponential coefficients. The characteristic equation of such systems has an infinite number of roots which cannot be found in the usual manner. Also, at times the only information available concerning a particular system may be frequency response data. For these reasons other methods have been developed for finding the time response of linear systems.

There are many different methods available for finding the time response, from frequency response data, of linear systems to unit impulse, unit step, or other arbitrary inputs. An excellent description and comparison of many of these methods is given by Donegan and Huss.[1]

The methods classified as "inverse transform methods" were found to be most effective in determining the response, to a unit impulse from frequency response data, of higher order systems. These methods are all based on the evaluation of the form of the inverse transform equation, equation 1, for positive values of time.

$$h(t) = \frac{2}{\pi} \int_0^\infty R_e[F(jw)] \cos twdw \qquad (1)$$

where:

$h(t)$ = time response of a system to a unit impulse
$F(jw)$ = frequency response of a system
$j = \sqrt{-1}$
t = time in seconds
w = angular frequency, radians per second

This equation has the disadvantage that if the system has an even-order pole at the origin, $R_e[F(jw)]$ will be infinite as w approaches zero. Also $R_e[F(jw)]$ in general will not approach zero quickly and, therefore, a large frequency range must be covered in the evaluation.

The accuracy of the Fourier method depends on the choice of a fundamental frequency w_f, which is quite difficult for high-order systems. The other methods either require previous knowledge of the form of the transfer function relating the system input and output or do not produce the required accuracy for higher order systems.

A graphical method, based on a different Fourier integral representation of the unit step function, has been developed by T. W. Sze[2] which will permit the finding of time response for systems containing delay functions. This method is also applicable to the finding of time response from frequency response data for any stable linear system. It has the particular attribute of eliminating the difficulties encountered with the Fourier and inverse Laplace transform methods for finding time response from frequency response data.

A different approach which produces the same end result as Sze's method may be found in a paper by A. Leonhard.[3]

Included in this paper will be two digital techniques for obtaining the transient response of linear control systems. The first is based on the conventional method of inverse Laplace transformation. The second is based on a Fourier integral

---

Paper 58-1291, recommended by the AIEE Computing Devices Committee and the Feedback Control Systems Committee of the Applications and Industry Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing September 11, 1958.

P. E. Lego is with the Westinghouse Electric Corporation, East Pittsburgh, Pa., and T. W. Sze is with the University of Pittsburgh, Pittsburgh, Pa.

This work is partly based on a Master of Science thesis by P. E. Lego at the University of Pittsburgh.

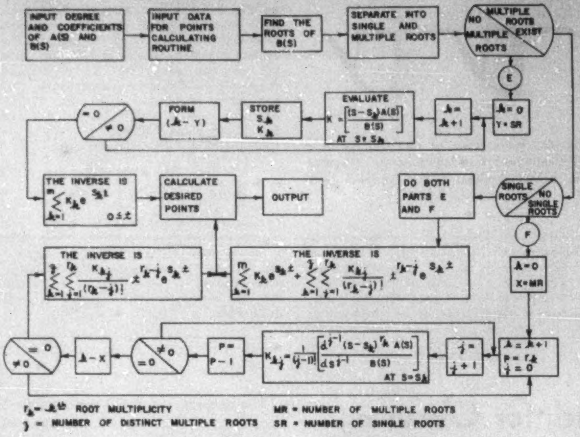

INPUT DEGREE AND COEFFICIENTS OF A(S) AND B(S)
INPUT DATA FOR POINTS CALCULATING ROUTINE
FIND THE ROOTS OF B(S)
SEPARATE INTO SINGLE AND MULTIPLE ROOTS
NO MULTIPLE ROOTS
MULTIPLE ROOTS EXIST
E
k = 0
Y = SR
k = k + 1
EVALUATE
K = [(S − S_k)A(S) / B(S)] AT S = S_k
STORE S_k K_k
FORM (k − Y)
= 0
≠ 0
THE INVERSE IS Σ_{k=1}^{m} K_k e^{S_k t} 0 ≤ t
CALCULATE DESIRED POINTS
OUTPUT
DO BOTH PARTS E AND F
SINGLE ROOTS
NO SINGLE ROOTS
F
k = 0
X = MR
THE INVERSE IS Σ_{k=1}^{m} K_k e^{S_k t} + Σ_{k=1}^{j} Σ_{j=1}^{r_k} K_{kj}/(r_k − j)! t^{r_k − j} e^{S_k t}
THE INVERSE IS Σ_{k=1}^{j} Σ_{j=1}^{r_k} K_{kj}/(r_k − j)! t^{r_k − j} e^{S_k t}
k = k + 1
P = r_k
j = 0
j = j + 1
K_{kj} = 1/(j − 1)! [d^{j−1}/ds^{j−1} (S − S_k)^{r_k} A(S)/B(S)] AT S = S_k
P = P − 1
≠ 0
= 0
k − X
= 0
≠ 0
r_k = kth ROOT MULTIPLICITY
j = NUMBER OF DISTINCT MULTIPLE ROOTS
MR = NUMBER OF MULTIPLE ROOTS
SR = NUMBER OF SINGLE ROOTS

Fig. 1. Flow diagram of inverse Laplace transform program

relation which uses frequency response data of the system.

Digital Computer Technique Based on Inverse Laplace Transformation

This section is concerned with systems whose behavior is known in the form of a system transfer function, which is a general rational algebraic proper fraction in the Laplace variable s, or in the form of a linear integro-differential equation from which the system transfer function may be derived.

In order to obtain the system time response, the process of inverse Laplace transformation must be performed on these transfer functions. This is a difficult, tedious procedure, especially for high-order systems.

A digital computer program has been developed, using the IBM (International Business Machines Corporation) *704* digital computer, which performs the inverse Laplace transformation of the following expression:

$$Q(s)=\frac{A(s)}{B(s)}=\frac{a_ms^m+a_{m-1}s^{m-1}+\ldots+a_1s+a_0}{b_ns^n+b_{n-1}s^{n-1}+\ldots+b_1s+b_0},\ (n>m) \tag{2}$$

This program contains the following restrictions:

1. The degree of the denominator, n, must be ≤ 30.
2. If there are multiple poles, the multiplicity of the pole is limited to four or less.

For computers of different varieties, these restrictions may be altered depending mainly on the memory capacity of the computer. One program with somewhat more stringent restrictions and applications was developed by Titus for use on a smaller computer.[4]

This program has the additional feature that it not only gives the residues and coefficients of all poles, but also calculates the magnitudes of system transient response at specified instants of time.

The logic of this program is presented in the flow diagram shown in Fig. 1. This logic is based on the conventional partial fraction expansion method for finding inverse Laplace transforms.

Digital Computer Technique Based on Fourier Integral Representation of Transient Response

In reference 1, the survey of different methods of obtaining system transient response placed these methods into several different categories and classified them as follows:

1. Inverse Laplace transformation methods:
   - (*a*). Floyd's method.
   - (*b*). Numerical integration method.
   - (*c*). Rectangular pulse method.
2. Fourier method.
3. Other methods:
   - (*a*). Schumaker's method.
   - (*b*). *P*-transform method.

In reviewing these methods, it has been found that they all require information on the system transfer function or system frequency reponse data. On the latter, the methods are mainly based on the integral relation shown in equation 1. The disadvantages of this integral have already been stated. A new relation for obtaining the transient response which overcomes these disadvantages will be shown here.

Theoretical Derivation

The general form of the Fourier integral in its exponential form is

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dw\int_{-\infty}^{\infty}e^{jw(t-u)}f(u)du \tag{3}$$

Let $F(p)$ be an operational function in which the differential operator $p=d/dt$. Thus $F(p)[f(t)]$ means that $F(p)$ operates on $f(t)$.

$$F(p)[f(t)]=\frac{1}{2\pi}\int_{-\infty}^{\infty}dw\int_{-\infty}^{\infty}F(jw)e^{jw(t-u)}\times f(u)du \tag{4}$$

If $f(t)$ is a unit step function, and $F(p)$ is the system transfer function in operational form, the transient response to a unit step input is $q(t)=F(p)[f(t)]$, which is the integral of equation 4. To simplify equation 4, assume that $F(jw)=R(w)+jI(w)$. Then

$$q(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dw\int_{-\infty}^{\infty}[R(w)\cos w(t-u)-I(w)\sin w(t-u)]f(u)du+j\frac{1}{2\pi}\int_{-\infty}^{\infty}dw\int_{-\infty}^{\infty}[I(w)\cos w(t-u)+R(w)\sin w(t-u)]f(u)du \tag{5}$$

Since $F(p)$ is a real operator for a physical system and $f(t)$ a real function, $q(t)$ should also be a real function. Thus the imaginary part of equation 5 must vanish.

$$q(t)=F(p)[f(t)]=\frac{1}{2\pi}\int_{-\infty}^{\infty}dw\int_{-\infty}^{\infty}\times[R(w)\cos w(t-u)-I(w)\sin w(t-u)]\times f(u)du \tag{6}$$

With the conventional representation, that $f(t)=1$ for $t>0$ and $f(t)=0$ for $t\leq 0$, the integral equation 3 does not exist since the integral of $f(t)$ over the infinite range becomes infinite. Here, the conventional representation of the unit step function is left and a new function $U(a,t)$ is named which has the following properties:

$$\lim_{a\to 0}U(a,t)=1 \quad \text{for } t>0$$
$$U(a,t)=0 \quad \text{for } t\leq 0$$
$$\lim_{t\to\infty}U(a,t)=0 \quad \text{for } a>0 \tag{7}$$

It is sufficient to select for this function

$U(a,t) = e^{-at}$ for $t > 0$. Substituting this function for $f(u)$ in equation 6,

$$F_a(p)[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} dw \int_{-\infty}^{\infty} \times$$

$$[R(w)\cos w(t-u) - I(w)\sin w(t-u)]e^{-au}du$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)w\sin tw}{a^2+w^2}\,dw + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{I(w)w\cos tw}{a^2+w^2}\,dw + \frac{a}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)\cos tw - I(w)\sin tw}{a^2+w^2}\,dw \quad (8)$$

Taking the limit as a approaches zero, equation 8 becomes $F(p)[f(t)]$ or $q(t)$:

$$q(t) = \lim_{a\to 0}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)w\sin tw}{a^2+w^2}\,dw + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{I(w)w\cos tw}{a^2+w^2}\,dw + \frac{a}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)\cos tw - I(w)\sin tw}{a^2+w^2}\,dw\right] \quad (9)$$

The evaluation of these limits yields:

$$\lim_{a\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)w\sin tw}{a^2+w^2}\,dw = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)\sin tw}{w}\,dw \quad (10)$$

$$\lim_{a\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{I(w)w\cos tw}{a^2+w^2}\,dw = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{I(w)\cos tw}{w}\,dw \quad (11)$$

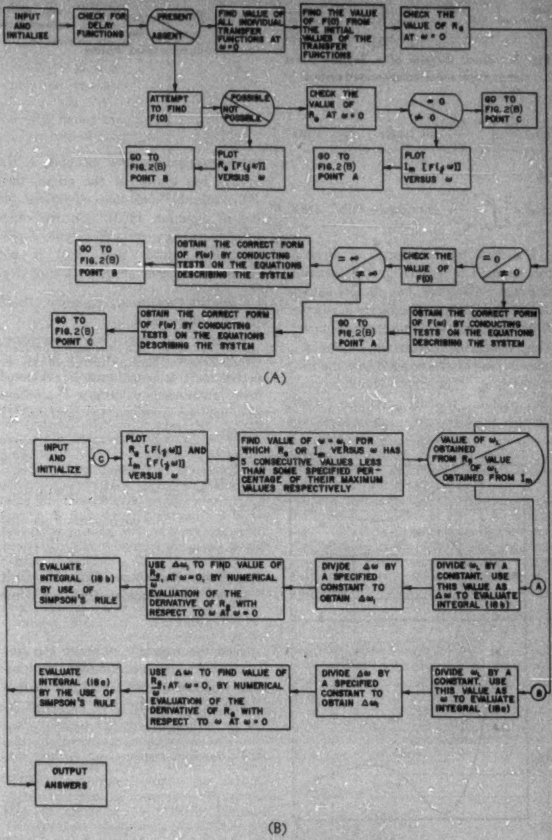
$$\lim_{a\to 0}\frac{a}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)\cos tw - I(w)\sin tw}{a^2+w^2}\,dw = \frac{1}{2}F(0) \quad (12)$$

The proof for equation 12 is shown in the Appendix. Thus

$$q(t) = F(p)[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{R(w)\sin tw}{w}\,dw + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{I(w)\cos tw}{w}\,dw + 1/2F(0) \quad (13)$$

Since $R(w)$ and $I(w)$ are the real and the imaginary part of the system transfer function $F(jw)$, $R(w)$ is then an even function and $I(w)$ is an odd function. Thus both $R(w)\sin tw/w$ and $I(w)\cos tw/w$ are even functions, and equation 13 may be rewritten as

$$q(t) = \frac{1}{\pi}\int_{0}^{\infty} \frac{R(w)\sin tw}{w}\,dw + \frac{1}{\pi}\int_{0}^{\infty} \frac{I(w)\cos tw}{w}\,dw + \frac{1}{2}F(0) \quad (14)$$


INPUT AND INITIALIZE
CHECK FOR DELAY FUNCTIONS
PRESENT
ABSENT
FIND VALUE OF ALL INDIVIDUAL TRANSFER FUNCTIONS AT ω = 0
FIND THE VALUE OF F(0) FROM THE INITIAL VALUES OF THE TRANSFER FUNCTIONS
CHECK THE VALUE OF Rₑ AT ω = 0
ATTEMPT TO FIND F(0)
POSSIBLE
NOT POSSIBLE
CHECK THE VALUE OF Rₑ AT ω = 0
= 0
≠ 0
GO TO FIG. 2(B) POINT C
GO TO FIG. 2(B) POINT B
PLOT Rₑ [F(jω)] VERSUS ω
GO TO FIG. 2(B) POINT A
PLOT Iₘ [F(jω)] VERSUS ω
GO TO FIG. 2(B) POINT B
OBTAIN THE CORRECT FORM OF F(ω) BY CONDUCTING TESTS ON THE EQUATIONS DESCRIBING THE SYSTEM
= ∞
≠ ∞
CHECK THE VALUE OF F(0)
= 0
≠ 0
GO TO FIG. 2(B) POINT C
OBTAIN THE CORRECT FORM OF F(ω) BY CONDUCTING TESTS ON THE EQUATIONS DESCRIBING THE SYSTEM
GO TO FIG. 2(B) POINT A
OBTAIN THE CORRECT FORM OF F(ω) BY CONDUCTING TESTS ON THE EQUATIONS DESCRIBING THE SYSTEM
INPUT AND INITIALIZE
C
PLOT Rₑ [F(jω)] AND Iₘ [F(jω)] VERSUS ω
FIND VALUE OF ω = ω₁ FOR WHICH Rₑ OR Iₘ VERSUS ω HAS 5 CONSECUTIVE VALUES LESS THAN SOME SPECIFIED PERCENTAGE OF THEIR MAXIMUM VALUES RESPECTIVELY
VALUE OF ω₁ OBTAINED FROM Rₑ
VALUE OF ω₁ OBTAINED FROM Iₘ
EVALUATE INTEGRAL (18b) BY USE OF SIMPSON'S RULE
USE Δω₁ TO FIND VALUE OF Rₑ/ω, AT ω = 0, BY NUMERICAL EVALUATION OF THE DERIVATIVE OF Rₑ WITH RESPECT TO ω AT ω = 0
DIVIDE Δω BY A SPECIFIED CONSTANT TO OBTAIN Δω₁
DIVIDE ω₁ BY A CONSTANT. USE THIS VALUE AS Δω TO EVALUATE INTEGRAL (18b)
A
EVALUATE INTEGRAL (18a) BY THE USE OF SIMPSON'S RULE
USE Δω₁ TO FIND VALUE OF Rₑ/ω, AT ω = 0, BY NUMERICAL EVALUATION OF THE DERIVATIVE OF Rₑ WITH RESPECT TO ω AT ω = 0
DIVIDE Δω BY A SPECIFIED CONSTANT TO OBTAIN Δω₁
DIVIDE ω₁ BY A CONSTANT. USE THIS VALUE AS ω TO EVALUATE INTEGRAL (18a)
B
OUTPUT ANSWERS

(A)

(B)

Fig. 2. Flow diagram of Fourier integral program

In a transient response study, the step function is applied at $t=0$. The system is assumed to be at rest prior to $t=0$, i.e.,

$$q(-t) = 0$$

or

$$-\frac{1}{\pi}\int_{0}^{\infty} \frac{R(w)\sin tw}{w}\,dw + \frac{1}{\pi}\int_{0}^{\infty} \frac{I(w)\cos tw}{w}\,dw + \frac{1}{2}F(0) = 0 \quad (15)$$

or

$$\frac{1}{\pi}\int_{0}^{\infty} \frac{R(w)\sin tw}{w}\,dw = \frac{1}{\pi}\int_{0}^{\infty} \frac{I(w)\cos tw}{w}\,dw + \frac{1}{2}F(0) \quad (16)$$

Here the final forms of transient response due to a unit step input relating to the frequency response data are written as

$$q(t) = \frac{2}{\pi}\int_{0}^{\infty} \frac{R(w)\sin tw}{w}\,dw \quad (17A)$$

$$q(t) = \frac{2}{\pi}\int_{0}^{\infty} \frac{I(w)\cos tw}{w}\,dw + F(0) \quad (17B)$$

In all physical systems, $F(jw)$ eventually approaches zero as the frequency increases. Thus $R(w)/w$ and $I(w)/w$ are both asymptotic to zero. Suppose that the values $R(w)/w$ and $I(w)/w$ become negligible when w reaches w_1 and w_2 respectively. The equations 17(A) and 17(B) may be written as

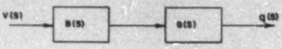


Fig. 3. Block diagram of a Taylor's series compensator series-compensated system

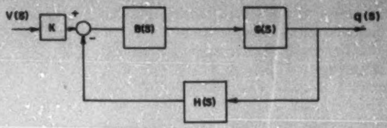


Fig. 5. Block diagram of a Taylor's series compensator feedback-compensated system

$$q(t_k)=\frac{2}{\pi}\int_0^{w_1}\frac{R(w)}{w}\sin t_k w\,dw \qquad \text{(18A)}$$

or

$$q(t_k)=\frac{2}{\pi}\int_0^{w_2}\frac{I(w)}{w}\cos t_k w\,dw+F(0) \qquad \text{(18B)}$$

Either of the two forms of integrals can be used in evaluating the transient response. However, two factors usually decide the choice. One factor is whether $R(w)/w$ or $I(w)/w$ approaches zero faster. This means to choose a smaller frequency range to be covered in computing the integral. The other factor depends on the system transfer function itself; whether the integral becomes nonexistent or indeterminant. For example, if the system has a pole at the origin, then $F(0)=$infinity, and that eliminates the use of equation 18(B). On the other hand, if the system is of a type of proportional control, then $R(0)$ is not equal to zero, and $\lim_{w\to 0} R(w)/w=$infinity. Under this condition, equation 18(A) should not be used.

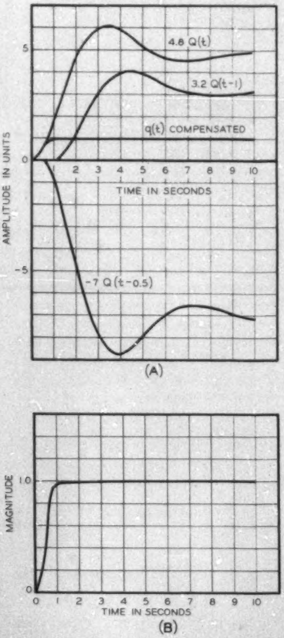


Fig. 4. Transient response to a unit step of Taylor's series compensator series-compensated second-order system

A—As given by inverse Laplace transform routine

B—As given by Fourier integral routine

In using the $I(w)/w$ curve, however, the value of $I(w)/w$ at $w=0$ is an indeterminant value and must be evaluated first. Its limiting value as w approaches zero can be obtained by applying L'Hospital's rule:

$$\lim_{w\to 0}\frac{I(w)}{w}=\lim_{w\to 0}\frac{\frac{d}{dw}I(w)}{1}=\frac{d}{dw}[I(w)]_{w=0}$$

$$=\text{initial slope of the } I(w)$$

curve versus w.

Transient Response Due to Unit Rate, Unit Acceleration, or Unit Impulse Input

Once the transient response $q(t)$ due to a unit step input is obtained, the transient response due to a unit rate or unit acceleration input may be evaluated by the use of the convolution integral. Thus

$r(t_k)=$transient response due to unit rate input

$$=\int_0^{t_k} q(t)dt \qquad (19)$$

$h(t_k)=$transient response due to unit acceleration input

$$=\int_0^{t_k} r(t)dt \qquad (20)$$

$i(t_k)=$transient response due to unit impulse input

$$=\left[\frac{d}{dt}\,q(t)\right]_{t=t_k} \qquad (21)$$

Computer Adaptation to the Preceding Fourier Method

A digital computer program has been developed for the IBM *704* computer based on the preceding method. The program is adapted to cases where the system transfer function is known and will compute the necessary frequency data. The program also handles those cases for which $R(w)$ and $I(w)$, or only frequency response data of the system, are known.

The logic of this program is shown by the flow diagrams of Figs. 2(A) and 2(B). Fig. 2(A), in conjunction with Fig. 2(B), shows the steps necessary when the system transfer function is the beginning point of the solution. Fig. 2(B), alone, shows the steps needed when frequency data are available to begin the problem.

Two examples are calculated to show the results of the computer programs.

EXAMPLE 1

This example was selected because it is possible to solve it by the conventional Laplace transform method as well as the Fourier integral method.

The system shown in block diagram form in Fig. 3 uses a delay line in series compensation. Using the data

$$B(s)=4.8-7e^{-0.5s}+3.2e^{-s} \qquad (22)$$

$$G(s)=\frac{1}{1+0.8s+s^2} \qquad (23)$$

with a unit step function as the input the system response function is

$$q(s)=\frac{4.8-7e^{-0.5s}+3.2e^{-s}}{s(1+0.8s+s^2)} \qquad (24)$$

The inverse Laplace transform gives $q(t)$, the transient response, as

$$q(t)4=.8Q(t)-7Q(t-0.5)+3.2Q(t-1)$$

where

$$Q(t)=1+1.091e^{-0.4t}\sin(0.916t-113.6\text{ degrees}) \qquad (25)$$

This response is found using the digital computer program for the inverse Laplace Transform and is plotted on Fig. 4(A).

The Fourier method calculates the response to a unit step function directly from the system transfer function given in equation 26.

$$F(s)=\frac{4.8-7e^{-0.5s}+3.2e^{-s}}{1+0.8s+s^2} \qquad (26)$$

The digital computer solution to this problem is shown in Fig. 4(B). A comparison of the results shows agreement within 0.5 per cent.

EXAMPLE 2

This method is selected so that the transient response cannot be found by

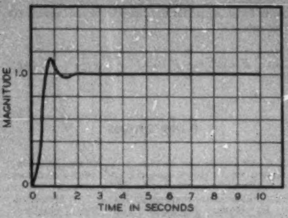
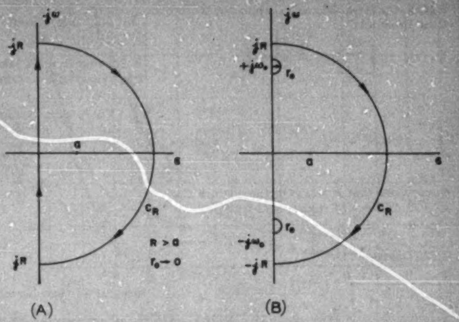


Fig. 6. Transient response to a unit step of a Taylor's series compensator feedback-compensated second-order system as given by Fourier integral routine

Fig. 7. Contour integration of analytic function $\psi(z)$

A—F(z) having no pole on imaginary axis

B—F(z) having conjugate pairs of poles on imaginary axis



the conventional inverse Laplace transform method since the number of poles is infinite. However, beginning with the system transfer function, or with frequency response data, the transient response may be found using the Fourier integral method described in this paper.

The block diagram of the system is shown in Fig. 5. Using the data

$$H(s)=0.4 \tag{27}$$

$$B(s)=4.8-7e^{-0.8s}+3.2e^{-s} \tag{28}$$

$$G(s)=\frac{1}{1+0.8s+s^2} \tag{29}$$

$$K=1.4 \tag{30}$$

the system transfer function may be expressed as

$$F(s)=\left[\frac{4.8-7e^{-0.8s}+3.2e^{-s}}{1+0.8s+s^2+(0.4)(4.8-7e^{-0.8s}+3.2e^{-s})}\right]1.4 \tag{31}$$

The unit step response of this system is calculated by the digital computer program directly from this transfer function. The result is shown in Fig. 6.

Conclusions

The adaptation of digital computers to the inverse Laplace transformation process and the Fourier integral method provides many advantages. The major advantages are as follows:

1. They offer a generalized technique for obtaining the transient response of a linear system.

(*a*). The digital program for the inverse Laplace transformation method, although limited to a system of 30th order or less with a multiplicity of poles of four or less, will solve most complex linear systems.

(*b*). The digital program for the Fourier integral method imposes no limit on the order of the system or the multiplicity of poles. Therefore, the program will handle any system so long as it is known to be linear and stable.

2. The Fourier integral method offers a process for obtaining the transient response of systems whose response cannot be found by the conventional Laplace transformation method. This method applies to the cases where only frequency response data of the the system is known, or where the system contains a characteristic equation which cannot be solved by conventional methods. This process, naturally, may be carried out without the use of a digital computer by using graphical integration and approximation.[2,3] It is, however, a tedious and time-consuming process, with a lower degree of accuracy, than a system designer would hesitate to attempt. With the digital computer in increasing use and availability, methods that usually require tremendous effort and time, may be carried out on the computer with ease. For instance, example 2 of this paper was carried out by graphical integration, using Simpson's rule, and by computer application. The former process took approximately 40 hours, using 20 intervals for integration. The latter took 3½ minutes, including input and output time, while using 100 intervals for integration by Simpson's rule.

3. The availability of these programs enables the designer to apply the Laplace transform technique to problems where the task of obtaining the Inverse Laplace Transform previously made this solution too difficult to undertake.[5]

Appendix. Proof of Equation 12

Equation 12 is stated as

$$\lim_{a\to 0}\frac{a}{2\pi}\int_{-\infty}^{\infty}\frac{R(w)\cos tw-I(w)\sin tw}{a^2+w^2}\,dw=\frac{1}{2}F(0) \tag{12}$$

This limit can be proved as follows.

The integral, equation 12, can be written

$$\frac{a}{2\pi}\int_{-\infty}^{\infty}\frac{R(w)\cos tw-I(w)\sin tw}{a^2+w^2}dw=\frac{a}{2\pi}\Re\left[\int_{-\infty}^{\infty}\frac{F(jw)e^{jtw}dw}{a^2+w^2}\right]=\frac{a}{2\pi}\Im\left[\int_{-j\infty}^{j\infty}\frac{F(jw)e^{jtw}d(jw)}{a^2-(jw)^2}\right] \tag{32}$$

This integral can be evaluated by a contour integration in the complex plane. Choose an analytic function $\psi(z)$ such that $\psi(z)=F(z)e^{tz}/a^2-z^2$, where $F(z)$ is known not to possess any pole on the right half side of the complex plane. Thus the only pole on the right half plane is at $z=a$. Integrate $\psi(z)$ in a clockwise direction around the boundary of the semicircle region, as shown in Figs. 7(A) and 7(B).

If $F(z)$ has no pole on the imaginary axis, the contour as shown in Fig. 7(A) will be used. Otherwise use the contour for integration as shown in Fig. 7(B). The semicircle has a radius R where $R>a$ so that the pole at $Z=a$ is enclosed. Thus by the Residue Theorem,

$$\oint\psi(z)dz=-2\pi jK_1 \tag{33}$$

where K_1 is the residue of $\psi(z)$ at the pole $z=a$, and the negative sign is due to the clockwise direction of contour integration. And

$$K_1=\lim_{z\to a}\frac{\dfrac{F(z)e^{tz}}{a^2-z^2}}{z-a}=\frac{-F(a)e^{at}}{2a} \tag{34}$$

Therefore

$$\oint\psi(z)dz=j\frac{\pi F(a)e^{at}}{a} \tag{35}$$

If $F(z)$ has poles on the imaginary axis, then from Fig. 7(B)

$$\oint\psi(z)dz=\int_{-jR}^{-j(w_0+r_0)}\psi(jw)d(jw)+\int_{-\pi/2}^{\pi/2}\psi(-jw_0+r_0e^{j\theta})d(r_0e^{j\theta})+\int_{-j(w_0-r_0)}^{j(w_0-r_0)}\psi(jw)d(jw)+\int_{-\pi/2}^{\pi/2}\psi(jw_0+r_0e^{j\theta})d(r_0e^{j\theta})+\int_{j(w_0+r_0)}^{jR}\psi(jw)d(jw)+\int_{C_R}\psi(z)dz \tag{36}$$

where C_R represents the right half of a circle with $z=R$, $(R>a)$. The limit of the integral, equation 36, as r_0 approaches zero then becomes:

$$\oint \psi(z)dz = \int_{-jR}^{jR} \psi(jw)d(jw) + \int_{C_R} \psi(z)dz \tag{37}$$

Thus

$$\int_{-jR}^{jR} \psi(jw)d(jw) = j\frac{\pi F(a)e^{at}}{a} - \int_{C_R} \psi(z)dz \tag{38}$$

everywhere that $R>a$. On the contour C_R, $|Z|=R$, or $Z=Re^{j\theta}$ with $-\pi/2 \leqq \theta \leqq \pi/2$,

$$|a^2-z^2| \leqq a^2-|z^2| = a^2-R^2$$

Consequently,

$$\left|\int_{C_R} \psi(z)dz\right| \leqq \int_{C_R} \frac{|F(z)|\,|e^{tz}|\,|dz|}{a^2-R^2}$$

$$= \int_{\pi/2}^{-\pi/2} \frac{|F(re^{j\theta})|\,|jRe^{j\theta}|\,|d\theta|}{a^2-R^2}$$

$$= \int_{\pi/2}^{-\pi/2} \frac{|F(Re^{j\theta})|\,Rd\theta}{a^2-R^2}$$

For a physical system with a transfer function of $F(jw)$, the limit of $F(Re^{j\theta})$ approaches zero as R approaches infinity. Also, since R has a higher power in the denominator than the numerator,

$$\lim_{R\to\infty}\left|\int_{C_R} \psi(z)dz\right| = 0 \tag{39}$$

From equation 38, taking the limit as $R\to\infty$, the following is obtained:

$$\int_{-j\infty}^{j\infty} \psi(jw)d(jw) = \lim_{R\to\infty}\int_{-jR}^{jR} \psi(jw)d(jw)$$

$$= j\frac{\pi F(a)e^{at}}{a} -$$

$$\lim_{R\to\infty}\int_{C_R} \psi(z)dz$$

$$= j\frac{\pi F(a)e^{at}}{a} \tag{40}$$

Integral 32 then becomes

$$\frac{a}{2\pi}\mathscr{J}\left[\int_{-j\infty}^{j\infty} \frac{F(jw)e^{jtw}d(jw)}{a^2-(jw)^2}\right]$$

$$= \frac{a}{2\pi}\mathscr{J}\left[\int_{-j\infty}^{j\infty} \psi(jw)d(jw)\right]$$

$$= \frac{a}{2\pi}\left[\frac{\pi F(a)e^{at}}{a}\right]$$

$$= \frac{1}{2}F(a)e^{at} \tag{41}$$

Now as a approaches zero the integral 40 becomes

$$\lim_{a\to 0}\frac{1}{2}F(a)e^{at} = \frac{1}{2}F(0) \tag{42}$$

References

1. Comparison of Several Methods for Obtaining the Time Response of Linear Systems to Either a Unit Impulse or Arbitrary Input from Frequency Response Data, J. J. Donegan, C. R. Huss. *Technical Note 3701*, National Advisory Committee for Aeronautics, Langley Aeronautical Laboratory, Langley Field, Va., Mar. 16, 1956.

2. Delay Line Compensator in Open and Closed Loop Systems, T. W. Sze. *Ph.D. Dissertation*, Northwestern University, Evanston, Ill., Aug. 1954.

3. Determination of Transient Response from Frequency Response, A. Leonhard. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., Nov. 1954.

4. A General Card Program for the Evaluation of the Inverse Laplace Transform, C. K. Titus. *Journal*, Association for Computing Machinery, New York, N. Y., vol. 2, no. 1, Jan. 1955, pp. 18–27.

5. The Influence of the Prestrike on Transmission-Line Lightning Performance, S. B. Griscom, J. W. Skooglund, A. R. Hileman. *AIEE Transactions*, vol. 77, pt. III, Dec. 1958, pp. 933–41.

Digital Calculation of Transient Performance of the Primary Coolant System in a Water Reactor

D. G. LEWIS
ASSOCIATE MEMBER AIEE

IN THE DESIGN of control systems for nuclear reactors, detailed knowledge of the rate of change of temperature in the reactor core is a very important factor. To evaluate this temperature as a function of time, it is necessary to determine both the heat added from the nuclear reaction itself, and the heat rejected through the cooling medium and through losses. The ability of the cooling medium to remove heat from the reactor is dependent upon the rate of coolant flow. This paper describes an application of a digital computer for computing coolant flow under various transient conditions which may arise during normal operation or during emergencies. The method used takes into account the following factors:

1. Inertia of the coolant fluid.
2. Inertia of rotating equipment (pumps and motors).
3. Frictional pressure drops in the flow paths.
4. Pump head flow characteristics.
5. Friction and windage losses in pumps and motors.
6. Transient effects in the rotor circuits of the pump motors.

Note that free convection of the fluid is not considered. Because of the heat transfer to and from the coolant, and because of the configuration of the flow paths, pressure heads are developed which result in a free convection component of flow. This component of flow becomes significant only when the total flow is less than 5% of normal flow. For flows greater than this, free convection may be safely ignored.

The flow paths in the coolant system are shown in Fig. 1. It is a 2-loop system as shown, so named because of the two parallel fluid paths, each with its own pump and driving motor. The flows in these two loops are not independent of each other, since the flow path through the reactor itself is common to both loops. The situation is analogous to that of a mutual impedance in a 2-loop electrical circuit, except that in the fluid case, the drop in potential (pressure) is not a linear function of the flow.

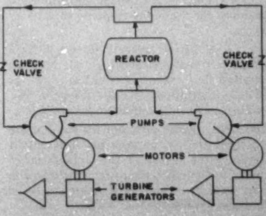


Fig. 1. Two-loop system arrangement

Paper 58-1314, recommended by the AIEE Computing Devices and Nucleonics Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 30, 1958; made available for printing September 11, 1958.

D. G. Lewis is with the General Electric Company, Schenectady, N. Y.

A portion of the material for this paper was obtained in the course of engineering development work under Contract No. *W-31-109 Eng-52*, sponsored by the General Electric Company which operates the Knolls Atomic Power Laboratory for the United States Atomic Energy Commission.

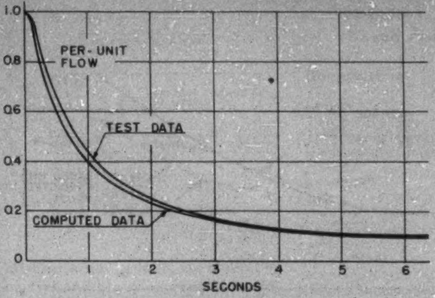


Fig. 2. Comparison of test data and computed results. Open-circuit on both pump motors at t=0. Note: Tests performed by Westinghouse Atomic Power Division for the Atomic Energy Commission

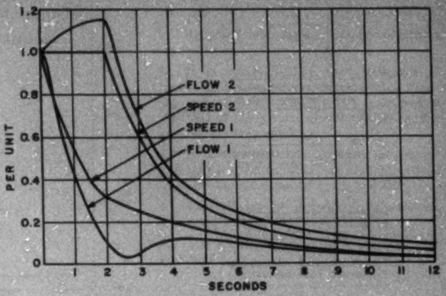


Fig. 3. Sequential failure. Motor 1 open-circuited at t=0, motor 2 open-circuited at t=2.0 seconds

In the system studied, the basic configurations of the two loops are identical. Thus the physical parameters of the two loops are the same, although the instantaneous flows and motor speeds may differ.

Energy to drive the coolant fluid is supplied by the pumps, which in turn are driven by induction motors supplied from individual generators. Each fluid loop has a check valve to prevent reverse flow, and it is also equipped with stop valves as required for control and maintenance.

In studying the flow in this system from the standpoint of protection against overheating of the reactor during emergencies, the important quantities to be determined are the total and loop flows and the motor speeds as functions of time. Quantities of secondary importance, but often needed in the design of protective arrangements, are the motor torques, currents, and rotor flux-linkages. To obtain these quantities, the differential equations for the system have been derived, and a digital computer has been used to develop particular solutions of these equations for cases of interest. The analytical approach and the numerical method for solving the differential equations are described in the section "Outline of Analysis and Computer Programming."

Transient Performance of Coolant System

Since the physical parameters of the two loops are the same, any operating condition or disturbance which affects both of the loops in exactly the same way can be analyzed much more easily than a disturbance which causes different conditions to exist in the two loops. The former case is referred to as a "symmetrical" case, the latter as a "nonsymmetrical" case. The equations and computer program for this study were developed to apply to nonsymmetrical cases, but can be used for symmetrical cases also.

The first set of results shown, Fig. 2, compares computed results with test data for a symmetrical case.[1] For this case, both loops are assumed to be operating in the steady state under rated conditions prior to the disturbance. At time zero, the pumps in both loops fail, due to sudden opening of the motor breakers. The curve shows the resulting decay of flow. The close agreement between the computed and test data indicates that both the differential equations and the numerical method used to solve them are a satisfactory approximation, despite the limitations inherent in the transient analysis of the electrical and fluid systems.

Fig. 3 shows both flows and speeds for both loops for a nonsymmetrical case. For this run, the initial conditions and the type of disturbance are the same as for the run of Fig. 2, except that the failure occurs in loop 1 at time zero and in loop 2 at $t=2.04$ seconds.

Fig. 4 shows flow, speed, and rotor flux linkage for another symmetrical case, namely, sudden loss of field on the generators supplying the pump motors. For this case, the motor currents and torques were also computed.

The computer program used to obtain these results has been used to study other cases as well, including loss of steam pressure on the turbine generators, 3-phase short circuits on the pump motors, motor shaft failure, and seizure of the pump impeller. Each of the cases requires some modification of the computer program, but the same basic program can be used for many different conditions as long as care is observed in making program changes correspond to changes in the physical system.

Outline of Analysis and Computer Programming

The general analysis of the system of Fig. 1 proceeds as follows: The equations of a multiloop fluid system are derived from the equations of Lagrange.[2] The force terms in these equations are then

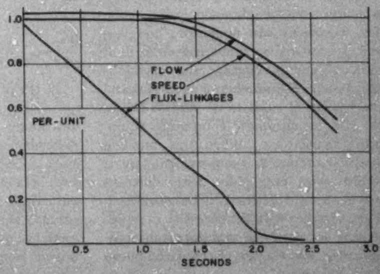


Fig. 4. Loss of excitation on generators

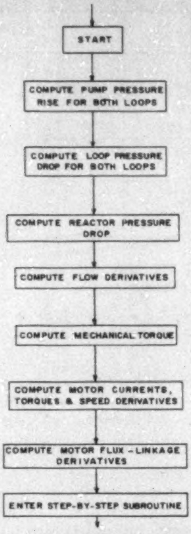

START
COMPUTE PUMP PRESSURE RISE FOR BOTH LOOPS
COMPUTE LOOP PRESSURE DROP FOR BOTH LOOPS
COMPUTE REACTOR PRESSURE DROP
COMPUTE FLOW DERIVATIVES
COMPUTE MECHANICAL TORQUE
COMPUTE MOTOR CURRENTS, TORQUES & SPEED DERIVATIVES
COMPUTE MOTOR FLUX-LINKAGE DERIVATIVES
ENTER STEP-BY-STEP SUBROUTINE

Fig. 5. Flow diagram of calculation of derivatives

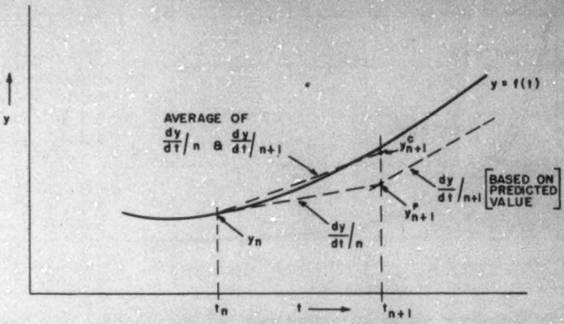


Fig. 6. Illustration of step-by-step method

related to the pump shaft torque by means of the steady-state pump head-flow characteristics. This shaft torque is obtained from the motor electrical torque less friction and windage terms. The motor torque can be calculated from motor current and flux linkages which are obtained in turn, by the 2-axis induction motor equations used previously to study the transient performance of induction motors.[3,4] These developments provide a set of four first-order differential equations for dependent variables in each loop of the system. Thus, in the 2-loop system, eight simultaneous first-order differential equations must be solved. The initial (steady-state) values of the dependent variables in these equations can be obtained quite easily. The equations are nonlinear, however, and a solution by numerical methods is required unless extensive simplifications are made. The number of equations and the variety of cases to be studied makes the use of automatic computing equipment mandatory. When this problem was first proposed, it was evident that it would arise repeatedly with different sets of physical constants to be considered. This, together with the necessity of representing accurately several important nonlinearities and the use of tabulated pump characteristics, led to the conclusion that a flexible digital computer program would be superior to an analog computer setup. Such a program was written for the International Business Machines Corporation *650* computer, and the results shown in Figs. 2 through 4 were obtained with this program. Complete results for symmetrical cases similar to those illustrated by Figs. 2 and 4 are obtained in about 20 minutes. For nonsymmetrical cases, such as the one illustrated by Fig. 3, complete results are obtained in about 40 minutes. A corresponding program has been written for the IBM *704* computer.

The computer program is divided into three main sections:

1. Calculation of derivatives.
2. Step-by-step numerical integration.
3. Output program.

The separation of the first two sections follows from the usual approach to numerical solutions of initial-value problems. Such a problem is expressible as a set of differential equations of the form

$$\begin{aligned}\frac{dy_1}{dt} &= f_1(t, y_1, y_2 \ldots y_n)\\ \frac{dy_2}{dt} &= f_2(t, y_1, y_2 \ldots y_n)\\ &\vdots\\ \frac{dy_n}{dt} &= f_n(t, y_1, y_2, \ldots y_n)\end{aligned} \qquad (1)$$

In the problem at hand, $n=8$, and the y's correspond to the flows, speeds, and motor flux linkages which appear as dependent variables in the system equations. The solution of such a problem can be obtained with several different numerical procedures of varying complexity and accuracy.[5,6] All of these methods, however, use the derivatives obtained from equation 1 together with values of the dependent variables known at a point t_k (or several points t_k, t_{k-1}, etc.) to obtain values of the variables at the point t_{k+1}. Repeated applications of this process, for which the term "step-by-step" is commonly used, give values of the dependent variables for any desired excursion of the independent variable, which in this problem is the time t.

To allow relative ease in making program modifications, the section of the computer program used for the calculation of derivatives was further subdivided. By keeping these subdivisions carefully identified, the chances for error in making program changes were substantially reduced. Fig. 5 illustrates the rather obvious nature of the these subdivisions and shows the sequence in which they occur as the derivative calculation is executed.

The step-by-step method used in the *650* program for this problem is known as the modified Euler's method and is often referred to as "slope extrapolation with trapezoidal improvement." It consists of the application of the two following equations.

Predictor (slope extrapolation):

$$y_j\Big|_{k+1}^{P} = y_j\Big|_{k}^{C} + \Delta t\left[f_j(t, y_1, y_2 \ldots y_j y_n)\Big|_{k}^{C}\right] \qquad \mathbf{(2)}$$

Corrector (trapezoidal improvement):

$$y_j\Big|_{k+1}^{C} = y_j\Big|_{k}^{C} + \frac{\Delta t}{2}\left[f_j(t, y_1 \ldots y_n)\Big|_{k}^{C} + f_j(t, y_1 \ldots y_n)\Big|_{k+1}^{P}\right] \qquad \mathbf{(2A)}$$

It can be seen that this corrector equation can be applied repeatedly at the point t_k until no further significant improvement can be obtained. However, since the evaluation of derivatives indicated by equations 1 must be made for each improvement, it is usually found most economical to make a single improvement as

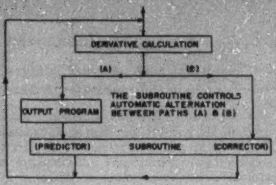


Fig. 7. Over-all flow diagram

in equation 2(A), reducing the integrating interval Δt if greater accuracy is needed. Fig. 6 illustrates the application of this method.

By separating the numerical integration section of the program from the other sections it is possible to write it in the form of a subroutine, with general features which allow it to be used for the solution of other sets of differential equations. Such a subroutine was developed in the work on this particular problem, and has been used successfully for other problems, such as the calculation of missile trajectories and the analysis of transient performance of groups of induction motors.[7]

The output program is made separate from the other sections for the following reasons:

1. It is not necessary to produce output data at each step in the numerical integration procedure.
2. Some of the desired output data is not normally computed during the evaluation of derivatives. These include magnitudes and angles of current and flux linkage which are normally computed in the form of rectangular components (α-β quantities).

Fig. 7 shows how the three sections of the program are used together to obtain the desired over-all operation.

Conclusions

A medium-sized digital computer, with a flexible program, provides a fast and economical way to study the transient performance of electromechanical-fluid systems if that performance can be defined by a set of ordinary first-order differential equations. Nonlinearities usually inherent in these systems can be handled readily by numerical methods. The necessity of considering a wide variety of transient disturbances means that the computer program must be modified frequently, but this can be done easily if care is taken to subdivide the program in correspondence with the actually separate parts of the physical system itself.

Appendix I. Nomenclature and Assumptions

Nomenclature

t = time
T = kinetic energy
U = potential energy
Q = generalized force
m = mass
v = velocity
l = length of flow path
A = area of flow path
g = acceleration of gravity
ρ = density
W = flow rate
ω = pump impeller and motor speed
ΔP_p = pump pressure rise
ΔP_L = frictional pressure drop in loop portion of flow path (not common to both loops)
ΔP_R = friction pressure drop in common portion of flow path (reactor)
$\Sigma l/A$ = geometry-dependent inertia factor for element of flow path
η_p = impeller efficiency
H_L = loop inertia factor
H_R = common path inertia factor
H_M = inertia factor for motor and impeller
T_M = mechanical torque
T_E = electrical torque
K_W = friction and windage loss coefficient
D = rotor diameter
L = rotor length
$\left.\begin{matrix}K\\A\\B\end{matrix}\right\}$ = constants defined when used
E = motor terminal voltage
I = motor armature current
ψ = motor rotor flux linkages
X = motor transient reactance: $X = L_s - M^2/L_r$
r = rotor resistance
r_M = resistance in magnetizing branch of motor equivalent circuit
M = mutual inductance
L_r = rotor self-inductance
L_s = stator self-inductance

Subscripts:

L = loop
R = reactor
$\left.\begin{matrix}1\\2\end{matrix}\right\}$ = loops 1 and 2
O = initial or rated value
N = normalized value
B = base quantity
$\left.\begin{matrix}\alpha\\\beta\end{matrix}\right\}$ = $\alpha-\beta$ axes used in motor analysis

Assumptions

Fluid System

1. System pressure drop is assumed to be a function of flow only.
2. Steady-state pump characteristics are assumed to hold even though the flow is transient.
3. It is assumed that the fluid is incompressible and that there is no loss of fluid from the system.
4. Thermal effects, both transient and steady-state, are neglected.
5. It is assumed that the constants of the two loops are the same.

Motors

1. Balanced rotor and stator windings are assumed.
2. It is assumed that the coefficient of mutual inductance between any stator winding and any rotor winding is a cosinusoidal function of the electrical angle between the axes of the two windings.
3. It is assumed that the rotor is smooth and that the self-inductances of all the windings are independent of rotor position.
4. The effects of saturation, hysteresis, and eddy currents are neglected.
5. The effects of transient stator currents are neglected.

Appendix II. Fluid System Equations

Development of General Equations

The dynamical equations which describe transient flow in a fluid system are easily derived, in the 1-dimensional case, by equating net power input to the rate of change of fluid kinetic energy. In the multiloop case, however, it is easier to apply the equations of Lagrange:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i}-\frac{\partial}{\partial q_i}(T-U)=Q_i \tag{3}$$

where T, U, and Q_i are respectively the kinetic energy, the potential energy, and the generalized forces acting on the system. The $\dot{q}_i$ are generalized velocities and the q_i are generalized displacements. Since the fluid is incompressible and body forces are not considered, the potential energy term may be omitted.

In a series of lengths of pipe, each of constant cross section, the kinetic energy may be expressed as follows:

$$KE=\sum\frac{1}{2}mv^2$$

$$=\frac{1}{2g}(W_1v_1^2+W_2v_2^2+\ldots)=\frac{1}{2g}\sum W_jv_j^2 \tag{4}$$

Now, since

$$W_j=A_jl_j\rho \tag{5}$$

and

$$v_j=\frac{W}{A_j\rho} \tag{6}$$

then

$$KE=\frac{W^2}{2g\rho}\sum l_j/A_j \tag{7}$$

In a multiloop system, the total kinetic energy is the sum of the fluid energies in the various flow paths.

The partial derivative of the kinetic energy with respect to displacement is zero; the derivative with respect to velocity is obtained as follows:

$$\frac{\partial}{\partial v_j}(KE)=\frac{\partial}{\partial v_j}\frac{(A_j\rho v_j)^2}{2g\rho}\sum l_j/A_j$$

$$=\frac{A_j^2\rho v_j}{g}\sum \Sigma l_j/A_j \tag{8}$$

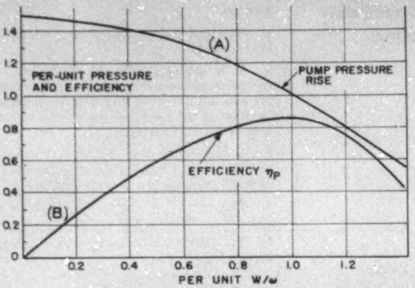


Fig. 8. Typical characteristics of single-suction radial flow centrifugal pump

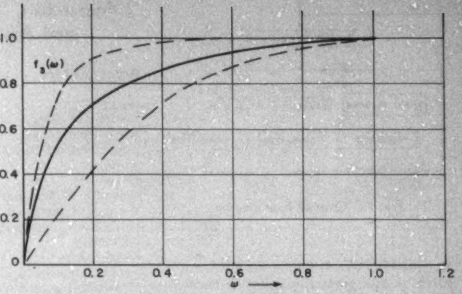


Fig. 9. Variation of design point impeller efficiency with speed

The force acting on the area A_j is simply $A_j\Delta P_i$, where ΔP_i is the net pressure rise around the flow path. The Lagrange equations then give, for a 2-loop system, the following:

$$\frac{A_j{}^2\rho}{g}\frac{dv_j}{dt}\sum l_j/A_{ji}=A_j\Delta P_i \qquad i=1,2$$

or

$$\frac{\sum l_1 A_{j1}}{g}\frac{dW_1}{dt}+\frac{\sum l_j/A_{jR}}{g}\frac{d(W_1+W_2)}{dt}=\Delta P_1 \tag{9}$$

$$\frac{\sum l_j/A_{j2}}{g}\frac{dW_2}{dt}+\frac{\sum l_j/A_{jR}}{g}\frac{d(W_1+W_2)}{dt}=\Delta P_2$$

in which the subscripts 1, 2, and R indicate respectively loop 1, loop 2, and the portion of the flow path common to both loops.

Table I. Frictional Loop Pressure Drop
Values in PU

Flow	Pressure Drop
0	0
0.04	0.0022
0.08	0.0083
0.12	0.0174
0.16	0.0287
0.20	0.0415
0.24	0.0594
0.28	0.0806
0.32	0.1049
0.36	0.1324
0.40	0.1631
0.44	0.1973
0.48	0.2344
0.52	0.2749
0.56	0.3170
0.60	0.3642
0.64	0.4132
0.68	0.4654
0.72	0.5211
0.76	0.5801
0.80	0.6425
0.84	0.7083
0.88	0.7757
0.92	0.8482
0.96	0.9224
1.00	1.0000
1.04	1.0809
1.08	1.1636
1.12	1.2513
1.16	1.3423
1.20	1.4351

Equations 9 now permit the calculation of the flow derivatives in terms of the pump pressure heads and the frictional pressure drops in the two loops.

Steam Pressure Rise

The net pressure rise is the difference between the pressure head produced by the pump and the frictional pressure drop caused by the instantaneous flow. The pressure head developed by the pump is developed from the following general affinity relations:

$$\frac{\Delta P_P}{\omega^2}=f_1\left(\frac{W}{\omega}\right) \tag{10}$$

The function f_1 is evaluated from a steady-state head-capacity characteristic at a known speed, and then ΔP_P can be formed for any flow and speed. A typical pump characteristic is shown in Fig. 8(A).

The frictional pressure drop is a function of flow only, with the major part generally of the form

$$\Delta P=K(W)^{\alpha} \tag{11}$$

The exponent α is commonly taken as 2, but may vary downward to 1.9 or lower, depending on system configuration. The results shown in this paper were obtained using a tabulated function shown in Table I.

A part of the frictional pressure drop may occur in a valve or other special element in the flow path, and this drop may be a function of flow more complicated than can be described by equation 11. In this study, the pressure drop in a check valve is included as part of the function in Table I.

It is important to recognize that, in a multiloop problem, components of the frictional pressure drop occur in portions of the flow path common to two or more loops. These drops must be evaluated separately, with the use of total flow in equation 11. A value $\alpha=2$ was used to calculate this component of pressure drop.

Impeller Efficiency

The mechanical torque acting on the pump motor may be found from the expression

$$T_M=\frac{\Delta P_P W}{\omega\eta_P}+K_W\omega^2 \tag{12}$$

which expresses impeller losses in terms of an efficiency η_P and includes rotor friction and windage as a function of speed. The efficiency is calculated as a product of two functions:

$$\eta_P=f_2\left(\frac{W}{\omega}\right)\ f_3(\omega) \tag{13}$$

The first function $f_2\ (W/\omega)$ expresses the relation between efficiency and relative flow rate. At a given speed, the efficiency is zero at zero flow, and reaches a maximum at some known value of flow. At design speed, the flow for maximum efficiency is the design point flow. This function is based on test data, and is shown in Fig. 8(B). Since test curves of this function on pumps of various designs do not vary appreciably, it may be considered reliable.

The second function, $f_3(\omega)$ expresses the reduction in maximum efficiency from the design point value at speeds other than rated speed. This function is also based on test data, but there is considerable variation in these data from one pump design to the next, so that this function should be selected with care. The function used in this study is shown in Fig. 9, with a typical range of variation shown in dashed lines. The evaluation of rotor friction and windage is discussed in Appendix III.

Appendix III. Motor Equations

Mechanical Losses

The motor friction and windage torque is assumed to be proportional to the square of speed, the constant of proportionality being derived from dimensional analysis and empirical data.

Fundamentally

$$\text{Power loss}=K\omega^3D^4L \tag{14}$$

For a given motor and pump, K can be determined by test and D and L are fixed; therefore equation 14 can be simplified:

$$\text{Power loss}=K^1\omega^3$$

and finally

$$T_{\text{windage}}=K_W\omega^2 \tag{15}$$

Currents, Torques, Flux Linkages, and Acceleration

The following equations, used for the calculation of motor quantities, are derived in reference 4, under the assumptions listed in Appendix I. The relations are the same for both motors, so the 1 and 2 subscripts are omitted.

$$I_\alpha=\frac{1}{X}\left(E_\beta-\frac{M}{L_r}\psi_\alpha\right) \tag{16}$$

$$I_\beta=\frac{1}{X}\left(-E_\alpha-\frac{M}{L_r}\psi_\beta\right) \tag{16A}$$

$$T_E=\frac{M}{L_r}(\psi_\alpha I_\beta-\psi_\beta I_\alpha) \tag{16B}$$

$$p\psi_\alpha=-\frac{r}{L_r}\psi_\alpha-\psi_\beta p\theta+\frac{r}{L_r}MI_\alpha \tag{16C}$$

$$p\psi_\beta=-\frac{r}{L_r}\psi_\beta+\psi_\alpha p\theta+\frac{r}{L_r}MI_\beta \tag{16D}$$

$$p\omega=\frac{1}{2H_M}(T_E-T_M) \tag{16E}$$

Equations 16 through 16(E) are general (though approximate) but must be modified for use in specific cases. For the case in which the motor breakers are assumed to open, zero current is forced by setting $1/X=0$ in equations 16 and 16(A). When this is done, the electrical torque computed from equation 16(B) will be zero also. With standard types of induction motors, during an open-circuit coastdown from rated conditions, the electrical torque developed would indeed be negligible, because the open-circuit core loss would be very small. However, with the "canned" motors used in reactor coolant pump applications, the rotor is coupled to a dissipative circuit even though the stator is open, and the power loss in this circuit must be supplied through an electrical torque which acts with the mechanical load to decelerate the rotor. The following equation is used to compute this torque for open-circuit cases.

$$T_{E(oc)}=-\frac{1}{R_M}\left(\frac{M}{L_r}\right)^2\omega(\psi_\alpha^2+\psi_\beta^2) \tag{17}$$

This equation can be derived easily from the common steady-state equivalent circuit of an induction motor.

In equations 16 through 16(E), E_α and E_β are the 2-phase components of applied voltage, and in normal operation $E_\alpha=1.0$, $E_\beta=0$. For the loss of excitation case, illustrated by Fig. 4, the constant value for E_α is replaced by a value computed from the following equation.[7]

$$E_\alpha=1.0\epsilon^{-t/T_{ds}'} \tag{18}$$

where

$$T_{ds}'=T_{do}'\frac{(X_d'+X_e)(X_q+X_e)+(r+r_e)^2}{(X_d+X_e)(X_q+X_e)+(r+r_e)^2} \tag{18A}$$

Reference 7 defines the quantities used in this equation.

Appendix IV. Complete System Equations

In this Appendix, the complete set of equations for the system will be listed. The flow diagram of Fig. 5 shows the sequence of calculations. The pump pressure rise is calculated with the use of the relationship given in equation 19, solved for ΔP_p:

$$\Delta P_p=\omega_N{}^2 f_1\left(\frac{W}{\omega}\right) \tag{19}$$

The function $f_1(W/\omega)$ is stored in the computer in tabular form. For use in this calculation, the impeller speed is normalized:

$$\omega_N=\omega/\omega_0 \tag{20}$$

where ω_0 is rated impeller speed.

Next, the loop pressure drops are computed:

$$\Delta P_{L1}=K_{L1}g_1(W_1) \tag{21}$$

$$\Delta P_{L2}=K_{L2}g_1(W_2) \tag{21A}$$

where $g_1(W)$ is the function given in Table I. Note that this function is normalized: $g_1(1.0)=1.0$. The constants K_{L1} and K_{L2} are introduced to give the correct per-unit loop pressure drop at unit flow.

The reactor pressure drop is computed from

$$\Delta P_R=K_R(W_1+W_2)^2 \tag{22}$$

The next step is the calculation of the flow derivatives. If, in equation 9, the terms of the form $(\Sigma l/A)/g$ are replaced by inertia factors $2H$, the results are

$$2H_{L1}\frac{dW_1}{dt}+2H_R\frac{d(W_1+W_2)}{dt}=\Delta P_1$$

$$2H_{L2}\frac{dW_2}{dt}+2H_r\frac{d(W_1+W_2)}{dt}=\Delta P_2$$

and, solving for the derivatives with $H_{L2}=H_{L1}=H_L$, yields

$$\begin{aligned}\frac{dW_1}{dt}&=A\Delta P_1-B\Delta P_2\\&=A(\Delta P_{P1}-\Delta P_{L1}-\Delta P_R)-\\&\qquad B(\Delta P_{P2}-\Delta P_{L2}-\Delta P_R)\end{aligned} \tag{23}$$

$$\begin{aligned}\frac{dW_2}{dt}&=A\Delta P_2-B\Delta P_1\\&=A(\Delta P_{P2}-\Delta P_{L2}-\Delta P_R)-\\&\qquad B(\Delta P_{P1}-\Delta P_{L1}-\Delta P_R)\end{aligned} \tag{23A}$$

in which

$$A=\frac{H_L+H_R}{2H_L(H_L+2H_R)} \tag{24}$$

$$B=\frac{H_R}{2H_L(H_L+2H_R)} \tag{24A}$$

Equations 23 and 23(A) are the ones used in the computer program to calculate flow derivatives.

The mechanical torque on the impeller shaft is computed next. Equations 12 and 13 are used; for T_{M1}

$$T_{M1}=\frac{(\Delta P_{P1})(W_1)}{(\omega_{N1})(\eta_{P1})}+K_W(\omega_{N1})^2 \tag{25}$$

$$\eta_{P1}=\left[f_2\left(\frac{W_1}{\omega_{N1}}\right)\right][f_3(\omega_{N1})] \tag{26}$$

The equations are the same for T_{M2} and η_{P2}.

The calculation of electrical quantities is next. The equations used are 16 through 16(E).

Appendix V. Calculation of System Constants

As is the case with many analytical studies, the determination of system constants for this problem is one of the most difficult phases of the study. There are two reasons for this. First, it is inherently difficult to obtain equipment performance data which will be valid for transient analysis. This is particularly true for the canned motors, which represent a departure from standardized designs. Second, in a composite system, the various items making up the system are supplied by different groups, and the analyst himself must often assume the responsibility for collecting data from several sources.

With respect to the difficulty of obtaining valid data, there is need to recognize that transient analyses of composite systems will be requested more and more frequently as control and protection arrangements become more complex. There appears to be a great need for standardized definitions of the quantities which become important in the study of transient motor phenomena.

In analyzing system problems, it is useful to express the important variables as decimal fractions of a specified reference value, in other words, in per unit (pu). When this is done, all the dimensional constants in the problem must be modified so that the system equations are satisfied when pu values are used throughout. Although this modification may require considerable calculation, it has advantages which make the preliminary work worth while. First, an evaluation of system performance in pu permits easy comparison with other systems of the same general type but of different sizes. An important consequence of this kind of comparison is the likelihood that errors will be detected, whether they occur in system parameters or in the results. A second important advantage is obtained if the calculations are made with automatic computing equipment, either analog or digital, for the expression of quantities in pu accomplishes all or nearly all of the scaling which is often required in the application of this equipment.

The first step in converting system data to pu is the selection of base quantities. It is common to select a power base, a time base, and a base value for one other quantity. These selections are sufficient to determine all other quantities; the selection of additional base quantities will lead to inconsistency. In this study of 2-loop flow transients, it was decided to use as base power the hydraulic power at rated flow, one second as base time, and rated flow as base flow. Since hydraulic power (HP)=(W)(ΔP), this selection also makes rated pressure drop correspond to base or unit pressure drop.

The evaluation of system constants in pu is shown in the following for a typical set of basic data. The base quantities are:

Base flow = rated loop flow $=3\times10^6$ pounds per hour
Base pressure drop = rated pump head = 130.6 psi (pounds per square inch)
Fluid density = 47.74 pounds/ft (per foot)

$$\text{Base power}=\frac{(W_{\text{base}})(\Delta P_{\text{base}})(144)}{(\rho)(3{,}600)(550)}\text{ HP}$$

$$=\frac{(3\times10^6)(130.6)(144)}{(47.74)(3600)(550)}=$$

$$586.67\text{ HP}$$

$$=437.48\ K_W$$

The system pressure drops at full flow are given as follows:

Loop (not common to both loops), psi	59.3
Reactor (common to both loops), psi	71.3
Total	130.6

The pressure drop coefficients are therefore

$$K_L=\frac{59.3}{130.6}=0.45406$$

$$K_R=\frac{71.3}{130.6}\times\frac{1}{4}=0.54594\times\frac{1}{4}=0.13648$$

Next must be calculated the inertia factors in the flow derivative equations. To do this, return to equation 9 and observe that the numerical exactness of this equation must be preserved when the flow and pressure drop are expressed in pu. To see what this implies, multiply the left-hand side by W_B/W_B (base flow) and divide each term in the equation by ΔP_B:

$$\frac{\left(\sum l/A_1\right)W_B}{(g)\Delta P_B}\frac{d(W_1/W_B)}{dt}+\frac{\left(\sum l/A_R\right)W_B}{(g)\Delta P_B}\frac{d(W_1+W_2)/W_B}{dt}=\frac{\Delta P_1}{\Delta P_B}$$

Flow and speed are now in pu. Now, consider the coefficient of the first flow derivative term. Multiply this coefficient by a unity ratio, as follows:

$$\frac{2\rho W_B}{2\rho W_B}\frac{\left(\sum l/A_1\right)W_B}{(g)\Delta P_B}=\frac{2\dfrac{\left(\sum l/A_1\right)W_B^2}{2g\rho}}{\dfrac{W_B\Delta P_B}{\rho}}$$

This manipulation, together with equation 7 and the expression for base power just given, show that this flow derivative coefficient is equal to twice the ratio of the loop fluid kinetic energy to the hydraulic power at rated flow, i.e.,

$$H_L=\frac{KE_{\text{loop}}}{\text{base power}}\text{ (in compatible units)}$$

Similarly, it can be shown that

$$H_R=\frac{KE_{\text{reactor}}}{\text{base power}}$$

These two relations make the calculation of the inertia coefficients simple, and provide a useful physical picture. The numerical calculation follows:

Fluid inertias (given)

$$\Sigma l/A_{\text{total/loop}}=200.5\text{ ft}^{-1}$$

$$\Sigma l/A_{\text{reactor}}=5\text{ ft}^{-1}$$

$$\Sigma l/A_{\text{loop}}=200.5-2\times5=190.5\text{ ft}^{-1}$$

Stored energy at full flow

$$KE_{\text{loop}}=\frac{(190.5)(3\times10^6)^2}{(2)(3600)^2(48.57)(32.2)(550)}$$

$$=76.8975\text{ hp-sec (seconds)}$$

$$KE_{\text{reactor}}=\frac{(5)(2\times3\times10^6)^2}{(2)(3600)^2(48.57)(32.2)(550)}$$

$$=8.07323\text{ hp-sec}$$

$$H_{\text{loop}}=\frac{76.8975}{586.67}=0.13107\text{ sec}$$

$$H_{\text{reactor}}=\frac{8.07323}{586.67}=0.013761\text{ sec}$$

These inertia factors have the dimensions of time. A useful physical check exists in the fact that a time equal to $2H$ is required for the inertia thus described to experience a change in speed (flow) equal to base speed (flow) when base force (pressure) is uniformly applied.

Now, A and B are as follows:

$$A=\frac{H_L+H_R}{2H_L(H_L+2H_R)}=3.4837\frac{1}{\text{sec}}$$

$$B=\frac{H_R}{2H_L(H_L+2H_R)}=0.33101\frac{1}{\text{sec}}$$

A similar analysis shows that the H_M of equation 16(E) can be calculated with the use of the kinetic energy of the impeller and rotor:

$$KE_M=\frac{1}{2}J\omega^2=\frac{(95.5)(377)^2}{(2)(32.2)(550)}$$

$$=383.211\text{ hp-sec}$$

In this case, the quantity given is the motor and impeller $WK^2=95.5$ pounds/ft². Then

$$H_M=\frac{383.211}{586.67}=0.65320\text{ sec}$$

It should be pointed out here that the pu representation of system inertias, using H-factors, makes clear the important fact that most of the system inertia resides in the rotors and impellers of the pumps. In such a system including both fluid and rigid-body inertia, the use of common physical units might obscure this point.

Next is calculated the constant K_W associated with rotor friction and windage loss. Preliminary calculations give the power loss due to friction and windage as

$$P=6.12\times10^{-4}\omega^3\text{ ft-pounds/sec}$$

where ω is given in radians/sec.

The pu power at rated speed is therefore

$$P=\frac{(6.12\times10^{-4})(377)^3}{(550)(586.67)}=0.10163\text{ per unit}$$

Since power equals the product of torque and speed, and rated speed = 1.0 in pu:

$$K_W=0.10163$$

The calculation of motor constants is usually made, and these constants are given, using the motor output rating as a pu base. These values must be modified to put them in pu on the hydraulic power base. The ratio of the bases can be found by calculating the motor power (or torque at unit speed) required to give unit hydraulic power.

$$(T_M)(\omega)=\frac{(W_B)(\Delta P_B)}{(\omega)(\eta_p)}\omega+K_W\omega^2\omega$$

$$=\frac{(1.0)(1.0)}{(1.0)(0.83)}(1.0)+(0.10163)(1.0)^2(1.0)$$

$$=1.306\text{ pu}$$

Therefore, the motor output should be 1.306 pu to give unit hydraulic power.

The motor constants used for this study were obtained partly from design data and partly on the basis of assumed similarity with other similar motors for which complete data were available. The values used are listed in Table II.

The numerical values used in the loss of excitation case are as follows:

$X_d=1.300$
$X_d'=0.210$
$X_q=0.900$
$r=0.005$ (approximately)
$T_{d0}'=3$ sec
$T_{dz}'=1.682$ sec
$X_e=0.456$
$r_e=0.989$

Note that X_e and r_e are calculated on the assumption of an average load of 0.9 pu kilovolt-amperes (on the generator base) at 0.89 power factor.

Appendix VI. Features of Numerical Integration Subroutine

As pointed out in the "Outline of Analysis and Computer Programming," the numerical manipulations required in the application of equations 2 and 2(A) can be programmed in the form of a subroutine for general use.

The accuracy obtainable in a solution using equations 2 and 2(A), or other such equations corresponding to other methods of numerical integration, is limited by the error in the calculation made at each step. In general, these components of error are cumulative. The error per step results from two causes: round-off error, which depends on the number of significant figures maintained in the computation, and truncation error, which results from the fact that the integration is performed over finite rather

Table II. Pump Motor Constants in PU

Constant	On Motor Base	On Hydraulic Power Base
x	0.1150	0.08804
M	2.775	2.1243
L_r	2.819	2.1584
r	0.00445	0.00341
R_M	13.06	10.00

than infinitesimal intervals. For a given set of integrating equations, the truncation error per step depends on the length of the integrating interval.

The length of the integrating interval therefore affects over-all accuracy in two ways. A change in this interval changes the truncation error per step, and also changes the number of steps required to complete a given solution, which affects the amount of accumulation of both types of error. Thus the selection of the integrating interval is an important decision with regard to the accuracy of the solution, and there is at least an indication that there may be an optimum interval for a given problem, in the sense that either increasing or decreasing the interval from this value will increase the over-all error. The selection of the interval is also an economic decision, since the time required to compute a step is the same whether the step is a long or a short one.

From the point of view of both over-all accuracy and economics, it is desirable to maintain uniform error per step throughout a solution, and this can be achieved only by allowing the interval to change during the process. If a numerical solution were to be obtained by hand, this control over the interval would be simple. If a digital computer is to be used, the control must be automatic; i.e., the decision to change the interval must be made by the computer.

The subroutine which was used to obtain the results shown in this paper incorporates an automatic estimate of the truncation error at each step, and it is able to make a change in the interval (doubling or halving) if the estimated error falls outside a specified range.

The operation of the error-estimating portion of the program is as follows. The program first compares the estimated error in a selected variable with the maximum allowable error. If the error is too large, the program goes back to the next previous set of *y*-values, halves the integrating interval, and tries again until the error is below the upper limit.

If the error is below the upper limit, the program next compares it with the minimum economic error. If it is greater than the minimum, the program goes ahead to the next interval. If the error is less than the minimum, the program goes ahead but it starts a counter as it does so. If the error is less than the minimum at each step for five successive steps, the calculation proceeds with a doubled integrating interval. The counter is returned to zero at every step unless the error is less than the minimum, so the count does not accumulate unless the low-error calculations occur in succession. The demand for five successive low-error calculations is arbitrary, and a lower or higher number can be used if it seems desirable.

References

1. RAPID FLOW TRANSIENTS IN CLOSED LOOPS, A. J. Arker, D. G. Lewis. *TID 7529*, pt. I, Book I, Atomic Energy Commission, Washington, D. C., 1957.

2. METHODS OF MATHEMATICAL PHYSICS, VOL. I (book), R. Courant, D. Hilbert. Interscience Publishers, Inc., New York, N. Y., 1953, p. 242 ff.

3. AN ANALYSIS OF THE INDUCTION MOTOR, H. C. Stanley. *AIEE Transactions*, vol. 57, 1938, pp. 751–55.

4. TRANSFER OF STEAM-ELECTRIC GENERATING-STATION AUXILIARY BUSSES, D. G. Lewis, W. D. Marsh. *Ibid.*, vol. 74, pt. III, June 1955, pp. 322–34.

5. NUMERICAL MATHEMATICAL ANALYSIS (book), J. B. Scarborough. Oxford University Press, London, England, 1950, p. 235 ff.

6. NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS (book), W. E. Milne. John Wiley & Sons, Inc., New York, N. Y., 1953.

7. SYNCHRONOUS MACHINES (book), C. Concordia. John Wiley & Sons, Inc., 1951, pp. 67–75.

Transistor Phase-Locked Oscillators

K. A. EDWARDS
NONMEMBER AIEE

O. GOLUBJATNIKOV
NONMEMBER AIEE

D. J. BRADY
NONMEMBER AIEE

THE USE of phase-locked oscillator control loops for the detection and tracking of signals in the presence of noise is well known and has been extensively analyzed.[1-4] Two completely transistorized phase-locked oscillator systems have been developed which perform these functions. System I operates at an input center frequency of 10.5 kc, and system II operates at 30 megacycles.

An additional function of frequency multiplication is obtained in the first of the two systems described by the addition of a frequency divider in the feedback path from the oscillator to the phase detector. This permits small input frequency shifts to be readily measured with a frequency counter. This basic system was suggested by Dr. L. J. Neelands.

The systems described have been designed for two different applications. In both systems, considerable advantage is obtained by use of transistors and semiconductor diodes. Some features peculiar to the systems described here, which result from the use of solid-state devices, are outlined as follows and will be described in detail later:

1. In order to obtain a narrow noise bandwidth, a long-time-constant integrating network is necessary. A simple *R-C* network is not suitable due to the lower impedance levels of transistor circuitry. A feedback integrator is used, and an analysis of the control loop stability including the effects of the integrator is shown.

2. The narrow bandwidth causes difficulty in locking on a signal during the search mode. This problem is solved with a nonlinear network in the integrator.

3. The phase detector permits use of the switching properties of transistors in a simple and reliable circuit. A signal is also provided for a novel phase-lock indicator.

4. A unijunction transistor provides a simple means of generating a search signal.

5. The freedom of transistors from microphonics and hum permits the elimination of a derivative feedback loop found necessary on comparable vacuum-tube equipment.

Linear Analysis

This analysis is based primarily on the transfer functions of system I; see Fig. 1. It includes the effects of the feedback frequency divider and feedback integrator. The equations apply to system II also; however, for system II, the frequency divider ratio M is unity and the feedback term $K_2(\tau_1-\tau_2)$ in the time constant τ_2 is zero.

The following simplifying assumptions are made in the analysis:

1. The phase detector output is a linear function of the phase difference between the feedback and reference inputs.

2. Leakage of the feedback capacitor and shunting effect of the amplifier input are neglected.

3. The amplifier output impedance is included in the time constant τ_2.

4. The oscillator control current and frequency are assumed to be linear functions of the amplifier feedback voltage.

5. Time constants of the oscillator control coil and the phase detector filter are neglected.

Paper 58-1276, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31 1958. Manuscript submitted April 9, 1958; made available for printing September 5, 1958.

K. A. EDWARDS and O. GOLUBJATNIKOV are with the General Electric Company, Syracuse, N. Y., and D. J. BRADY is with United Engineers, Inc., Boston, Mass.

Credit is due D. L. Hagen of the General Electric Advanced Engineering Program for his contributions to the analysis of system I and for the development of the lock indicator and search circuit. J. T. Holland, Jr., suggested the method of injecting the search signal in system I. Dr. L. J. Neelands gave assistance with determining the system requirements and with the analysis. F. J. Ryan and W. R. McKinley contributed much experimental work and refinements of the circuits.

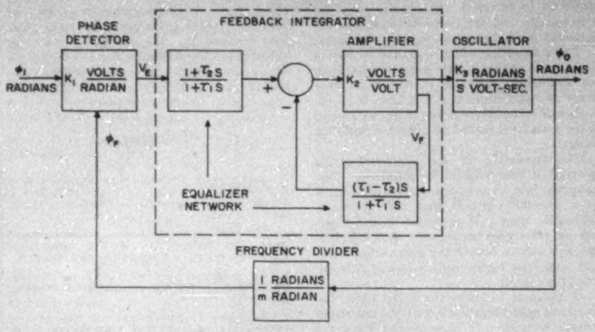


Fig. 1. Transfer functions of system I

Referring to Fig. 1, the transfer function of the feedback integrator v_F/v_E as a function of the Laplace operator S is

$$\frac{V_F}{V_E}(S)=\frac{K_2(1+\tau_2 S)}{1+[K_2(\tau_1-\tau_2)+\tau_1]S} \tag{1}$$

where τ terms are time constants and K terms are gains.

The open-loop gain of the phase lock system is

$$G(S)=\frac{K_1K_2K_3}{M}\left\{\frac{1+\tau_2 S}{1+[K_2(\tau_1-\tau_2)+\tau_1]S}\right\}\left\{\frac{1}{S}\right\} \tag{2}$$

where M is the frequency division ratio.

The closed-loop transfer function is

$$\frac{\phi_0}{\phi_1}(S)=\frac{M\left(\frac{K}{T_3}\right)(1+\tau_2 S)}{S^2+\left(\frac{1+K\tau_2}{\tau_3}\right)S+\frac{K}{\tau_3}} \tag{3}$$

where ϕ terms are phase angles in radians and

$$\tau_3=K_2(\tau_1-\tau_2)+\tau_1$$

$$K=\frac{K_1K_2K_3}{M}$$

From servomechanism convention, equation 3 can be written,[5]

$$\frac{\phi_0}{\phi_1}(S)=\frac{M\left[\omega_n^2+\left(2\zeta\omega_n-\frac{\omega_n^2}{K}\right)S\right]}{S^2+2\zeta\omega_n S+\omega_n^2} \tag{4}$$

where

ζ = ratio of actual to critical damping
ω_n = resonant frequency of loop

It is seen that

$$\omega_n=\sqrt{\frac{K}{\tau_3}} \tag{5}$$

$$2\zeta\omega_n=\frac{1+K\tau_2}{\tau_3} \tag{6}$$

$$\frac{2\zeta}{\omega_n}=\frac{1}{K}+\tau_2$$

$$\zeta=\frac{\omega_n}{2K}+\frac{\tau_2\omega_n}{2}$$

For most cases,

$$\tau_2\omega_n>>\frac{\omega_n}{K}$$

$$\zeta\approx\frac{\tau_2\omega_n}{2} \tag{7}$$

NOISE BANDWIDTH

The noise bandwidth may be obtained by evaluating the integral,

$$B=\int_{-\infty}^{+\infty}\left|\frac{\phi_0}{\phi_1}(j\omega)\right|^2 d\omega \tag{8}$$

Equation 4 in terms of $j\omega$ becomes

$$\frac{\phi_0}{\phi_1}(j\omega)=\frac{1+j2\zeta\frac{\omega}{\omega_n}\left(1-\frac{\omega_n}{2\zeta K}\right)}{1+j2\zeta\frac{\omega}{\omega_n}-\left(\frac{\omega}{\omega_n}\right)^2}M \tag{9}$$

Substituting equation 9 in 8,

$$B=M^2\omega_n\times$$

$$\int_{-\infty}^{+\infty}\frac{1+4\zeta^2\left(\frac{\omega}{\omega_n}\right)^2\left(1-\frac{\omega_n}{2\zeta K}\right)}{1-(2-4\zeta^2)\left(\frac{\omega}{\omega_n}\right)^2-\left(\frac{\omega}{\omega_n}\right)^4}d\frac{\omega}{\omega_n} \tag{10}$$

The integration of an equation of this form was done by Gruen[1] with the following result:

$$B=M^2\pi\omega_n\frac{4\zeta^2-4\zeta\frac{\omega_n}{K}+\left(\frac{\omega_n}{K}\right)^2+1}{2\zeta} \tag{11}$$

For

$$\frac{\omega_n}{K}\rightarrow 0$$

and in terms of input bandwidth,

$$\frac{B}{M^2}=\pi\omega_n\frac{4\zeta^2+1}{2\zeta} \tag{12}$$

Ingham[8] has shown that B is minimum for $\zeta=0.5$. Fig. 2 shows a calculated curve of B/M^2 as a function of ζ.

HOLD-IN RANGE

Gruen has analyzed hold-in range for a linear system where the phase detector output is proportional to the sine of the phase difference.[1] In the system under consideration, the phase detector output is directly proportional to the phase difference over a wide range. An analysis of linear effects follows.

In order to keep the equations in terms of the input frequency, ϕ_0/M rather than ϕ_0 will be used.

$$\phi=\phi_1-\frac{\phi_0}{M} \tag{13}$$

From equation 4,

$$\frac{\phi_0}{M}=\frac{\omega_n^2+\left(2\zeta\omega_n-\frac{\omega_n^2}{K}\right)S}{S^2+2\zeta\omega_n S+\omega_n^2}\phi_1 \tag{4}$$

Combining equations 4 and 13

$$(S^2+2\zeta\omega_n S+\omega_n^2)\phi=\left(S^2+\frac{\omega_n^2}{K}S\right)\phi_1 \tag{14}$$

If the difference between the input and feedback frequency $\Delta\omega=s\phi_1$ is a constant, equation 14 becomes

$$(S^2+2\zeta\omega_n S+\omega_n^2)\phi=\frac{\omega_n^2}{K}\Delta\omega \tag{15}$$

The steady-state solution is

$$\phi=\frac{\Delta\omega}{K} \tag{16}$$

$$\therefore\ \Delta\omega_{\text{hold-in}}=K\phi_{\max} \tag{17}$$

The maximum excursion of ϕ is determined by the saturation of the amplifier and oscillator control characteristics once the phase detector characteristic is known.

Description of System I

Fig. 3 shows a block diagram of one system which operates as follows: Assume

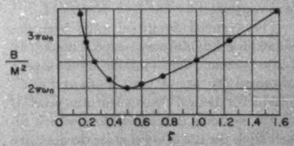


Fig. 2. Input noise bandwidth versus damping factor

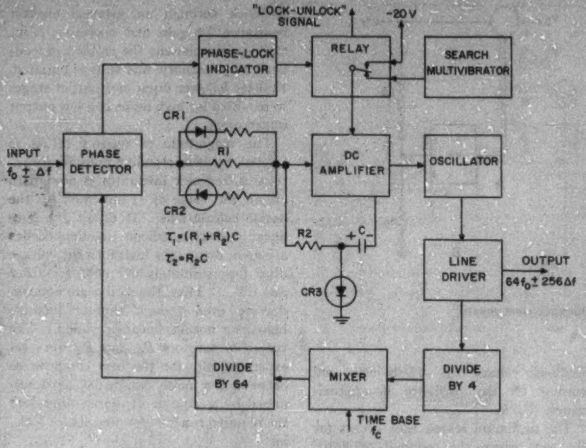

"LOCK-UNLOCK" SIGNAL
-20V
PHASE-LOCK INDICATOR
RELAY
SEARCH MULTIVIBRATOR
CR1
R1
CR2
INPUT $f_0 \pm \Delta f$
PHASE DETECTOR
DC AMPLIFIER
OSCILLATOR
$\tau_1=(R_1+R_2)C$
$\tau_2=R_2C$
R2
+C−
CR3
LINE DRIVER
OUTPUT $64f_0 \pm 256\Delta f$
DIVIDE BY 64
MIXER
DIVIDE BY 4
TIME BASE f_c

Fig. 3. Block diagram of system I

that initially the system is not phase-locked. A d-c signal from the phase detector causes the lock indicator to de-energize the relay thus giving a no-lock indication and connecting the search multivibrator to the output stage of the d-c amplifier.

As the multivibrator waveform swings positive, the oscillator frequency goes to its lowest value in less than 1 second. The rapid retrace is due to *CR-3* switching to the conducting state, permitting rapid discharge of capacitor *C*. As the multivibrator swings negative, *CR-3* ceases to conduct, and the waveform is integrated to produce a sawtooth of current in the oscillator control coil which sweeps the oscillator frequency slowly toward its upper frequency limit.

The bandwidth of the system is increased during the search and pull-in by the conduction of *CR-1* or *CR-2* which switches shunting resistance across R_1 to reduce the integrator time constant.

When the oscillator frequency reaches a value such that the frequency divider output equals $f_0 \pm \Delta f$, the system pulls into synchronism. The phase detector output then goes to its locked-in range of voltage. This causes the lock indicator to energize the relay, disconnecting the search multivibrator, and returning the d-c amplifier to the power supply. The oscillator output is then $64f_0 \pm 256\Delta f$.

The clock frequency is synchronized with the input frequency f_0. When incoming intelligence is received, f_0 is shifted by Δf, but the clock frequency does not shift. Thus, the Δf is multiplied by 256, but f_0 is only multiplied by 64. The magnitude of Δf is proportional to the magnitude of the intelligence.

The frequency multiplication feature permits the Δf to be determined to a greater accuracy in a shorter sampling time using a frequency counter.

Calculation of System I Constants

The following performance for the phase-locked loop was set as a goal:

1. Noise bandwidth ≤ 50 cps (cycles per second) or 314 rad/sec (radians per second).
2. Search and pull-in time = <5 sec.
3. Hold-in range = 1,000 cps or 6,280 rad/sec.
4. Divider ratio $M = 256$.

The open-loop gain may be calculated from

$$K=\frac{\Delta\omega_{\text{hold-in}}}{\phi_{\max}} \qquad (17)$$

$\phi_{\max}$ is set at 0.5 rad to insure small-signal operation of the phase detector.

$$K=12{,}560 \text{ per sec} \qquad (18)$$

The loop resonant frequency may be calculated from

$$\omega_n=\left(\frac{B}{\pi M^2}\right)\left(\frac{2\zeta}{4\zeta^2+1}\right) \qquad (12)$$

Ingham[3] has shown that the maximum pull-range for a given noise bandwidth occurs at $\zeta=\sqrt{3}/2$. However, to insure stability with age and parameter variations, the damping factor ζ is made unity.

$$\omega_n=40 \text{ rad/sec} \qquad (19)$$

The time constant τ_1 is

$$\tau_1=\frac{K}{\omega_n^2} \qquad (5)$$

For $\omega_n \leq 40$ rad/sec

$$\tau_1 \geq 7.9 \text{ sec} \qquad (20)$$

The time constant τ_2 is

$$\tau_2 \approx \frac{2\zeta}{\omega_n} \qquad (7)$$

$$\tau_2 \approx 0.05 \text{ sec} \qquad (21)$$

Analysis of the Nonlinear Integrator

The nonlinear integrator accomplishes the following; see Fig. 4:

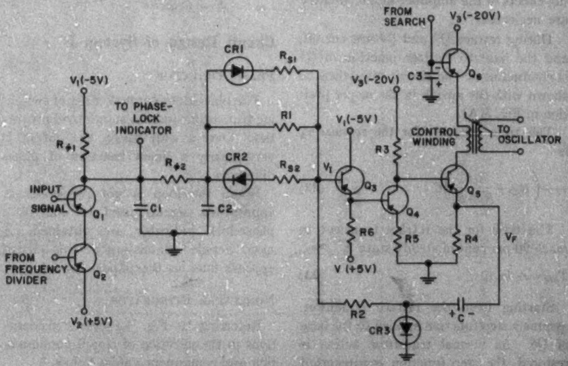

$V_1(-5V)$
TO PHASE-LOCK INDICATOR
CR1
R_{S1}
R1
$R_{\phi 1}$
$R_{\phi 2}$
CR2
R_{S2}
V_1
INPUT SIGNAL
Q_1
C1
C2
FROM FREQUENCY DIVIDER
Q_2
$V_2(+5V)$
FROM SEARCH
$V_3(-20V)$
C3
Q_6
$V_3(-20V)$
$V_1(-5V)$
CONTROL WINDING
TO OSCILLATOR
R3
Q_3
Q_4
Q_5
R6
R5
R4
V_F
V (+5V)
R2
+C−
CR3

Fig. 4. Nonlinear integrator

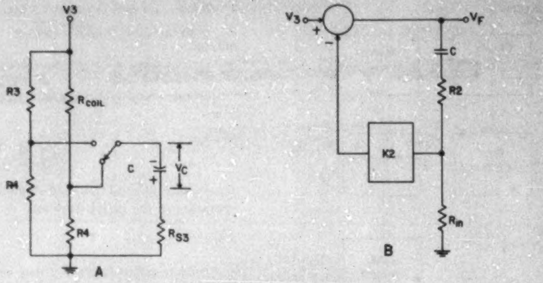


Fig. 5. Nonlinear integrator equivalent circuits

1. When the system is phase-locked, all diodes are nonconducting and linear analysis applies.
2. During the retrace of the search multivibrator (positive swing), *CR-3* conducts and discharges capacitor *C*.
3. During the search (negative swing) diode *CR-1* conducts shunting R_1 with R_{s1}, thus increasing the resonant frequency ω_n of the feedback loop. With a signal present, the phase detector output is positive, thus transistor Q_4 is cut off.
4. When the system is locked, sudden changes in phase detector output, if of great enough magnitude, will produce a drop across R_1 due to the charging current of capacitor *C*, sufficient to produce conduction of silicon junction diodes *CR-1* or *CR-2*. The bandwidth and damping factor are both increased, thus reducing the chance of unlocking due to spurious disturbances. The lock-in range for a given search time also increases.

The retrace time may be calculated from the equivalent circuit of Fig. 5(A). The diodes and transistors involved are assumed to be perfect switches, and shunting effects of the amplifier input circuitry are neglected.

During retrace *Q3* and *Q4* are cut off, and the base to emitter junction of *Q5* is conducting. The circuit may then be shown with the switch in the upper position in Fig. 5(A).

The time constant for the retrace τ_R is

$$\tau_R = \left(R_{S3} + \frac{R_3 R_4}{R_3 + R_4} \right) C \tag{22}$$

The time for the retrace transient to reach 90 per cent of steady state is

$$T_{Rm} = \tau_R \ln 10 \tag{23}$$

Starting from the retrace condition, assume a step function applied to the base of *Q6*. As normal transistor action is restored, the step function is integrated to produce as exponential increase of the feedback voltage V_F. The control coil current for the oscillator is approximately V_F/R_4.

The maximum search time occurs for the case where the required oscillator frequency, and hence V_F is at its highest value. The equivalent circuit of Fig. 5(B) is used to calculate the maximum search time T_{sm}. Resistance R_1 is shunted by R_{s1} since the phase detector voltage is always positive, causing *CR-1* to conduct, when the system is not phase-locked, and a signal is present. Using ordinary linear transient analysis T_{sm} is

$$T_{SM} = -\tau_B \ln \left(\frac{1 - V_f/V_3}{1 - \tau_A/\tau_B} \right) \tag{24}$$

where

$\tau_A = (R_{in} + R_2)C$

$\tau_B = [R_2 + R_{in}(1 + K_2)]C$

$R_{in} = R_{\phi 2} + \dfrac{R_1 R_{s1}}{R_1 + R_{s1}}$

$R_{2\phi}$ = approximate internal resistance of the phase detector

Circuit Design of System I

PHASE DETECTOR

The phase detector makes use of switching transistors and has some novel properties. Over a wide range, the output is very nearly a linear function of phase difference.

When the loop is not phase-locked, appropriate signals used to control a phase-lock indicator are obtained. A more detailed discussion of this circuit appears later in the paper.

NONLINEAR INTEGRATOR

Referring to Fig. 4, some considerations in the selection of circuit configuration and component values follow.

A single-stage voltage amplifier with feedback through an external emitter resistance for gain and operating point stability was chosen; the problem of feedback loop stability was thus eliminated. Emitter follower input and output stages were added for high input and low output impedances.

The magnitude of resistor of R_1 is limited by bias stability considerations; thus, a feedback integrator is needed to obtain the value of τ_3 indicated by the design calculations. To obtain the nonlinear operation, silicon junction diodes are used, due to the high forward voltage drop (approximately 0.5 volt) for *CR-1* and *CR-2*. Thus, the diodes are nonconducting even though slightly forward biased for normal linear operation. The values of resistors R_{s1} and R_{s2} were set experimentally for the best compromise between low noise bandwidth as determined by ω_n and short search time over the required range of frequencies for lock-in.

PHASE-LOCK INDICATOR

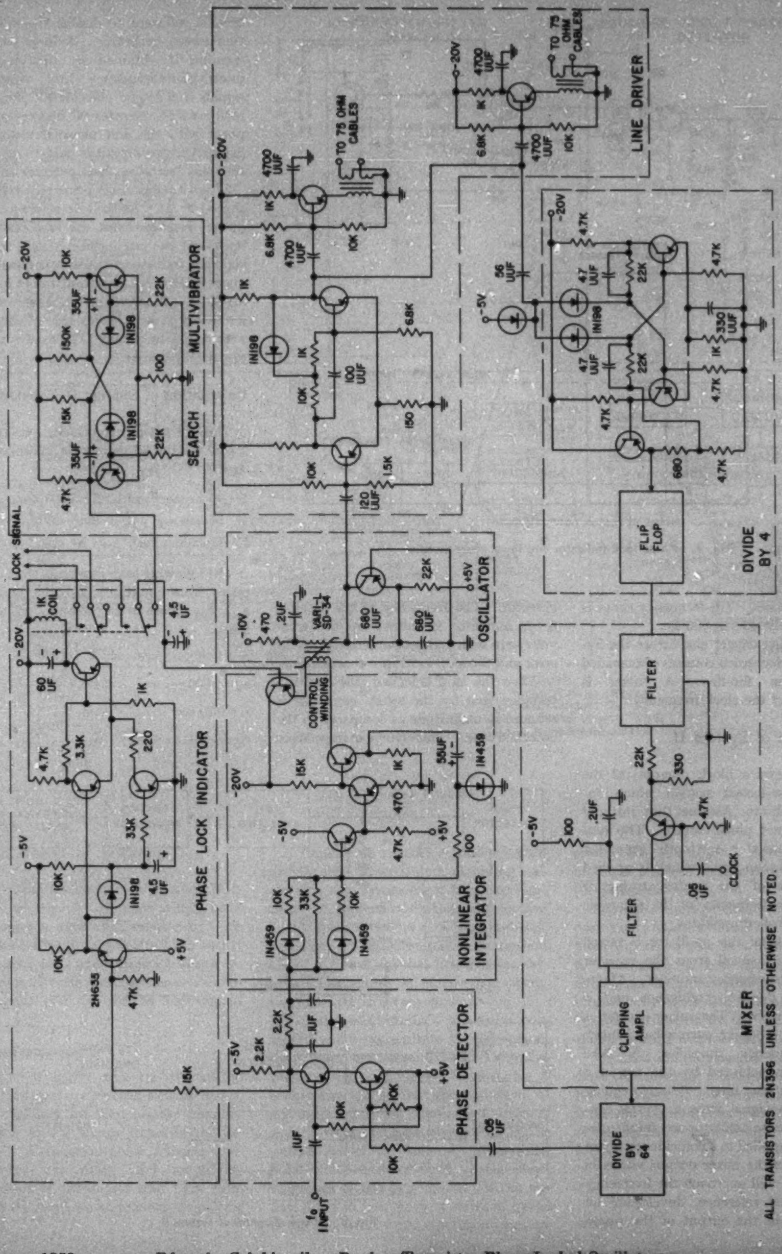
The phase detector properties are an important part of the operation of the phase-lock indicator. The phase detector d-c output voltage must be considered for three possible cases in the lock indicator operation; refer to Figs. 6 and 7. If the system is phase-locked, the phase detector voltage is in zone *A* (linear operation). The unlock indications are from zones *B* and *C*. The voltage in zone *B* only occurs when one of the signals is missing. If both signals appear, but the system is unlocked, the phase detector voltage jumps to the lower limit of zone *C*.

During unlock, the search multivibrator is connected to the nonlinear integrator; see Fig. 6. The search multivibrator generates an unsymmetrical square wave which gives a retrace time of approximately 0.4 sec and a search time of approximately 3 sec. The diodes limit the reverse voltage swing on the transistor bases to a safe value.

OSCILLATOR AND LINE DRIVER

The oscillator is a Colpitts circuit with frequency control through variation of the direct current in the control winding.

Since the required output from the system is a square wave, the oscillator drives a Schmitt trigger which supplies a square wave. The Schmitt trigger output drives the frequency divider and two emitter followers coupled through pulse transformers to four 75-ohm coaxial cables. The square wave output pulse to each of the cables is 2.5 volts peak-to-peak amplitude with a rise time of less than


SEARCH MULTIVIBRATOR
PHASE LOCK INDICATOR
LOCK SIGNAL
1K COIL
PHASE DETECTOR
NONLINEAR INTEGRATOR
OSCILLATOR
VARI-L SD-34
CONTROL WINDING
LINE DRIVER
TO 75 OHM CABLES
DIVIDE BY 4
FLIP FLOP
FILTER
MIXER
CLIPPING AMPL
DIVIDE BY 64
CLOCK
f0 INPUT
2N635
IN198
IN459
-20V
-10V
-5V
+5V

ALL TRANSISTORS 2N396 UNLESS OTHERWISE NOTED.

Fig. 6. System I schematic.

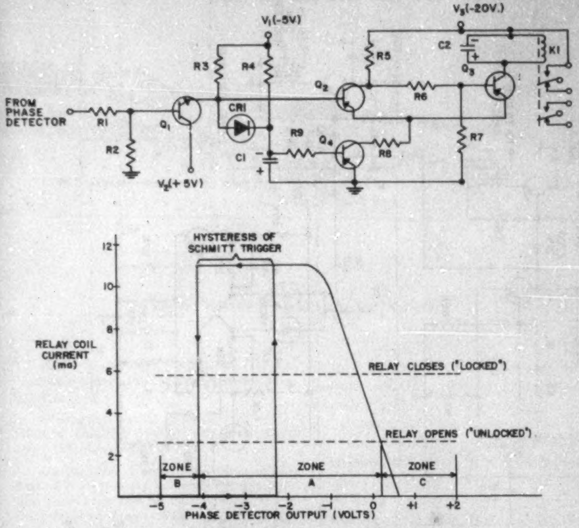

V₁(-5V)
V₃(-20V.)
R3
R4
R5
C2
K1
Q₃
Q₂
R6
FROM PHASE DETECTOR
R1
Q₁
CR1
R9
Q₄
R7
R2
C1
R8
V₂(+5V)
HYSTERESIS OF SCHMITT TRIGGER
RELAY COIL CURRENT (ma)
12
10
8
6
4
2
RELAY CLOSES ("LOCKED")
RELAY OPENS ("UNLOCKED")
ZONE B
ZONE A
ZONE C
-5
-4
-3
-2
-1
0
+1
+2
PHASE DETECTOR OUTPUT (VOLTS)

Fig. 7. Phase-lock indicator and input characteristic

0.2 microsecond. The frequency range is approximately 400 to 900 kc.

The Schmitt trigger also drives the frequency divider which consists of cascaded Eccles-Jordan flip-flops. A mixer is used to inject the clock frequency.

Description of System II

Fig. 8 shows a block diagram of the second phase-locked system which operates as follows: Assume that initially the loop is not phase-locked. The unijunction transistor sawtooth generator produces a sweep signal which is applied at the base of the control transistor sweeping the frequency of the saturable reactor controlled oscillator.

The output of this oscillator is mixed with a $f_0 \pm \Delta f$ signal from the receiver intermediate-frequency amplifier. Δf represents the desired intelligence and is masked by noise. The mixer output is amplified in a 455-kc narrow-band intermediate-frequency amplifier. A 6-kc bandwidth is achieved by the use of a magnetostriction filter. The foregoing signal and a signal from a crystal-controlled time-base oscillator are amplitude-limited and applied to a coincidence phase detector. As the mixer output varies in frequency, it will approach the frequency of the crystal reference, developing an error signal in the output of the phase detector. The frequency of the nominal f_0+f_c oscillator is shifted by the error voltage in such a direction as to reduce the error and the loop locks into synchronism.

When the loop is locked, the load impedance seen by the sweep generator is reduced in magnitude as compared to the unlocked case. This change in impedance level is sufficient to disable the unijunction sweep generator. A loop lock indication is obtained by applying the intermediate-frequency and time-base signals to a second coincidence phase detector which operates 90 degrees out of phase with the first phase detector. A Schmitt trigger circuit is used to activate the relay from the phase detector output.

When intelligence is being received f_0 is shifted by Δf. As the loop references frequency remains fixed, the loop readjusts itself and the intelligence Δf will add or subtract from the oscillator nominal frequency f_0+f_c. The output mixer following the oscillator strips the signal carrier frequency f_0 and leaves the intelligence with a bias frequency f_c to resolve directional ambiguities.

Calculation of System II Constants

The following performance was set as a goal for the 30-megacycle phase-locked loop:

1. Noise bandwidth, $B=5{,}000$ cps.
2. Search and pull-in time <5 sec.
3. Hold-in range, $\Delta\omega=10^6$ cps.

The following loop constants were computed from equations 5, 7, 12, and 17. ϕ_{max} was chosen as 0.5 rad and ζ was made unity.

$$K=1.26\times10^7/\text{sec} \tag{17}$$

$$\omega_n=4{,}000 \text{ rad/sec} \tag{12}$$

$$\tau_1=0.79 \text{ sec} \tag{5}$$

$$\tau_2=5\times10^{-4} \text{ sec} \tag{7}$$

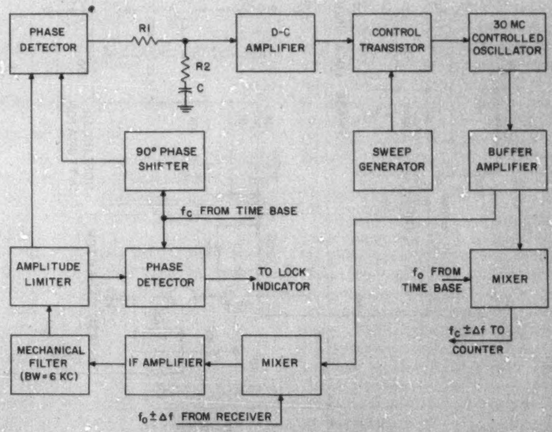

PHASE DETECTOR
R1
R2
C
D-C AMPLIFIER
CONTROL TRANSISTOR
30 MC CONTROLLED OSCILLATOR
90° PHASE SHIFTER
SWEEP GENERATOR
BUFFER AMPLIFIER
fc FROM TIME BASE
AMPLITUDE LIMITER
PHASE DETECTOR
TO LOCK INDICATOR
f0 FROM TIME BASE
MIXER
fc ±Δf TO COUNTER
MECHANICAL FILTER (BW=6 KC)
IF AMPLIFIER
MIXER
f0 ±Δf FROM RECEIVER

Fig. 8. Block diagram of system II

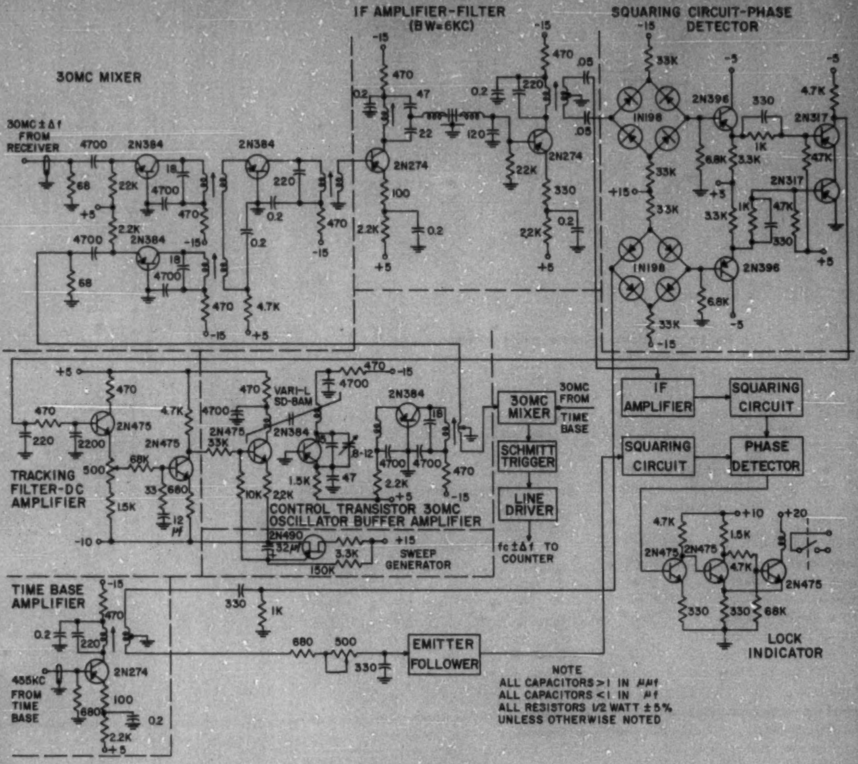

IF AMPLIFIER-FILTER (BW=6KC)
SQUARING CIRCUIT-PHASE DETECTOR
30MC MIXER
30MC ± Δf FROM RECEIVER
2N384
2N274
2N396
2N317
IN198
TRACKING FILTER-DC AMPLIFIER
2N475
VARI-L SD-8AM
CONTROL TRANSISTOR 30MC OSCILLATOR BUFFER AMPLIFIER
2N490
SWEEP GENERATOR
30MC MIXER
30MC FROM TIME BASE
SCHMITT TRIGGER
LINE DRIVER
fc ± Δf TO COUNTER
IF AMPLIFIER
SQUARING CIRCUIT
SQUARING CIRCUIT
PHASE DETECTOR
LOCK INDICATOR
TIME BASE AMPLIFIER
455KC FROM TIME BASE
EMITTER FOLLOWER
NOTE
ALL CAPACITORS >1 IN μμf
ALL CAPACITORS <1 IN μf
ALL RESISTORS 1/2 WATT ± 5%
UNLESS OTHERWISE NOTED

Fig. 9. System II schematic

Circuit Design of System II

The circuit schematic of system II is shown in Fig. 9. Circuits which have been used in more than one location are repeated in block form. Coincident-phase-detector and unijunction-sweep-generator circuits are considered novel in this application and will be discussed in detail.

Phase Detector

Phase detectors may be constructed with only passive elements, such as diodes, but amplifiers are then usually required to make up for energy losses in the detector. A transistor phase detector combines the switching and amplifying properties of transistors and thus eliminates the need for additional amplifying stages.

A number of transistor-phase-detector circuits were investigated, including unijunction and symmetrical transistor[4] circuits. A coincidence phase detector was finally selected because of its simplicity, immunity to signal amplitude variations above a threshold, and linear characteristic. The circuit schematic and operating characteristics are shown in Fig 10.

Transistors Q_1 and Q_2 are series connected and both saturate if, and only if, both bases are simultaneously negative with respect to ground. If capacitor C is disconnected so that the waveform at the output may be examined, a series of square waves will be observed. If the signals are of the same frequency, the pulse width is determined by the phase difference of the two inputs. If the signals differ in frequency, a series of pulses of varying widths and repetition rates are produced. The addition of capacitor C produces a direct voltage proportional to the phase difference.

Transistor storage effects encountered during saturation seriously limit the maximum repetition rate at which the phase detector can be driven. Linearity of this circuit can be improved and driving rates increased by controlled saturation or preventing saturation by the use of a diode clamp or using nonlinear feedback.

The function of capacitor C_1 is to overdrive the transistor momentarily. Knowing the value of the stored base charge from the manufacturer's data, the value of the capacitor can be computed from the formula:

$$C_1 = \frac{Q_s}{E_1} \tag{25}$$

where

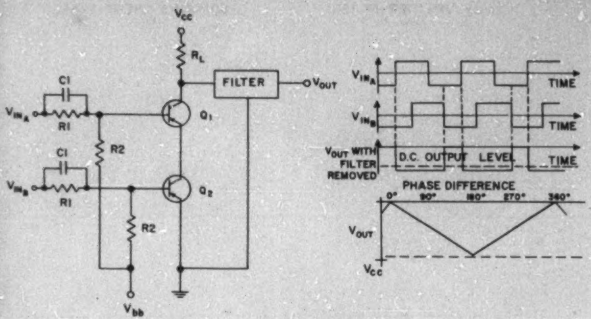


Fig. 10. Phase detector and operating characteristic

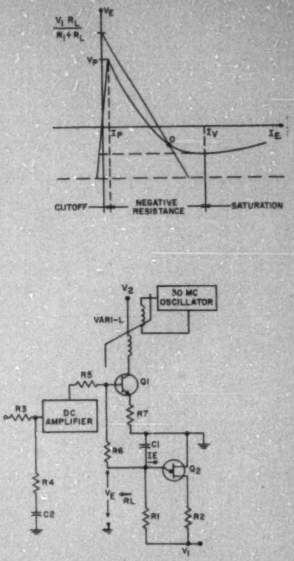


Fig. 11. Unijunction transistor sweep generator

C_1 = coupling capacitance, micromicrofarads
Q_S = stored base charge, micromicrocoulombs
E_s = input voltage, volts

The same result can be obtained experimentally by adjusting C_1 for minimum storage and fall time without an overshoot.

The value of R_1 is determined by the lower limit of h_{fe} of the transistors to be used. Bias voltage V_{bb} and resistor R_2 should be chosen so that the base potential will be slightly positive at the maximum temperature of operation.

There is a small delay due to the time it takes for the base to fall to ground potential. This delay is approximately equal for both channels resulting in a very small net effect. As the I_{co} of the transistor increases with temperature, the base becomes less positively biased, reducing the rise time. This effect is compensated by the decrease of the alpha-cutoff frequency due to decreased mobility of minority carriers.

SWEEP CIRCUIT

The negative-resistance characteristic of the unijunction transistor is utilized conveniently in a relaxation sawtooth oscillator to generate the necessary search signal. The circuit is shown in Fig. 11.

As the loop locks, the equivalent load resistance R_L seen by the sweep circuit is reduced in magnitude. The circuit must be designed so that it is monostable when the loop locks and is astable when the loop is unlocked. In terms of the operating characteristic, the load line of the R_1 and R_L parallel combination must intersect the V_p axis above the peak point V_p to assure that the diode will begin conducting during search. The load line must intersect below V_p when the loop is locked.

The intersection at point O must be in the negative resistance portion of the characteristic to ensure that the circuit begins to recharge the capacitor and does not operate at a stable point on the saturation portion of the curve. In the following expressions, it is assumed that $(R_L)NL > (R_L)L$ and that the base of Q_1 is approximately at ground potential.

During Search:

$$\left|\frac{V_1(R_L)NL}{R_1+(R_L)NL}\right| \geq |V_p| \tag{26}$$

and

$$\left|\frac{R_1(R_L)NL}{R_1+(R_L)NL}\right| \geq |R_{min}| \approx \left|\frac{V_p-V_v}{I_p-I_v}\right| \tag{27}$$

During Lock:

$$\left|\frac{V_1(R_L)L}{R_1+(R_L)L}\right| < |V_p| \tag{28}$$

Experimental Results

Table I shows experimental results. Calculated values where applicable were obtained from the relations derived earlier.

SYSTEM I

For determining the maximum Δf over which the system searched and locked on, the time base frequency f_e and the input frequency f_0 were set so that the mixer output was in the center of the pass band of the mixer output filter. The −3-decibel bandwidth of this filter is 150 kc, and the −30-decibel bandwidth is 230 kc with a center frequency of 672 kc. Thus, the frequency range is limited by the operating range of the d-c amplifier. The Δf realized is greater than the design objective which shows that the d-c amplifier has adequate operating range.

The design objective for the resonant frequency ω_n is determined by the desired noise bandwidth. A lower limit is im-

Table I. Experimental Results

Design Objective	Measured Data	Calculated Results	Remarks
	System I		
$\Delta f \geq \pm 500$ cps	$\Delta f = \pm 700$ cps		
$\omega_n < 40$ rad/sec	$\tau_1 = 40$ sec $K = 10{,}600$/sec	$\omega_n = 16$ rad/sec	$f_0 - \Delta f \approx 10$ kc
	$K = 23{,}200$/sec	$\omega_n = 24$ rad/sec	$f_0 \approx 10.5$ kc
	$K = 14{,}100$/sec	$\omega_n = 19$ rad/sec	$f_0 + \Delta f \approx 11$ kc
$T_{rm} + T_{sm} < 5$ sec	$T_{rm} \approx 0.4$ sec $T_{sm} \approx 3.0$ sec		
	System II		
$\Delta f \geq \pm 500$ kc	$\Delta f > \pm 500$ kc	Δf dependent on intermediate-frequency amplifier bandwidth and search rate	
$\omega_n < 4{,}000$ rad/sec	$\omega_n = 3{,}800$ rad/sec		
Search time < 5 sec	Search time = 2 sec		

posed by the maximum search time. The values of ω_n obtained indicate that the specifications have been met. The time constant τ_3 of the feedback integrator was measured with a Tektronix *535* oscilloscope. Phase difference measurements were made with the same oscilloscope and frequency measurements with a frequency counter. These measurements were made closed loop from which loop gains were determined at three different operating points. The variations in loop gain (and in ω_n) appear to be due to nonlinearity of the oscillator control coil in the lower range of oscillator frequency (upper range of $f_0 \pm \Delta f$) and to nonlinearity of the input current versus base to emitter voltage for the d-c amplifier transistors at the upper range of oscillator frequency.

The use of the nonlinear diode network permits the search time to be less than the design objective.

Since the phase detector gives limiting action, the theoretical performance of this system in the presence of random noise can be expected to follow the bandpass limiter case discussed by Jaffe and Rechtin.[7]

SYSTEM II

System II, with minor modifications, has been used to track signals in the 10- to 40-megacycle range. The sweep circuit described permits a Δf greater than ±500 kc. The ω_n with the circuit constants shown in Fig. 9 was measured at approximately 3,800 rad/sec which correlates well with the calculated value. The search rate is limited primarily by the passband characteristics of the 455-kc intermediate-frequency amplifier. The search time with the circuit constants shown is approximately 2 sec. At higher search rates, the lock-in reliability is reduced. Signal-to-noise ratios of approximately -10 decibels at the input from the receiver have been achieved.

Conclusions

The applicability of transistors to phase-locked oscillator control loops has been successfully demonstrated wth two systems operating in widely separated portions of the frequency spectrum. Also, the transistor circuitry is amenable to analysis. The use of nonlinear properties of semiconductor diodes and transistors permits seemingly conflicting objectives to be attained.

References

1. Theory of A. F. C. Synchronization, W. J. Gruen. *Proceedings*, Institute of Radio Engineers, New York, N. Y., 1956, pp. 1045–46.

2. Closed-Loop Automatic Phase Control, P. F. Ordung, J. E. Gibson, B. J. Shinn. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 375–81.

3. The Design of an A. P. C. Synchronizing Loop, W. E. Ingham. *Report no. RW/8*, Electronic Musical Instrument Research Laboratories Ltd., Hayes, Middlesex, England, Apr. 1956.

4. The Pull-In Range of an A. P. C. Loop, E. L. C. White. *Report no. RK/94*, Electronic Musical Instrument Research Laboratories, Nov. 1955.

5. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., New York, N. Y., 1948.

6. A Symmetrical Transistor Phase Detector for Horizontal Synchronization. *Report no. LB-914*, Radio Corporation of America, New York, N. Y., June 1953, pp. 1–8.

7. Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance over a Wide Range of Input Signal and Noise Levels, R. Jaffe, E. Rechtin. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. IT-1, Mar. 1955, pp. 66–76.

8. How to Design Transistorized Schmitt Triggers, R. B. Hurley. *Electronic Equipment*, New York, N. Y., Aug. 1956.

1957 Magnetic Amplifier Bibliography

AIEE COMMITTEE REPORT

THIS bibliography covers domestic and foreign literature for the year 1957. Some articles from the year 1956, not previously listed, are also included. The patents listed cover those issued between July 1956 and December 1957. Earlier magnetic amplifier literature is covered in a Magnetic Amplifier Applications Subcommittee report covering the literature for the years 1951 through 1956.[16]

The bibliography is divided into four major subdivisions as follows:

I. English Language Articles
II. Foreign Language Articles
III. Domestic Patents
IV. Foreign Patents

Within each major subdivision, papers and patents are classified according to subject matter as follows:

A. *Theory, Analysis, and Design*
Papers dealing primarily with the theory of operation of magnetic amplifiers, or with the analysis and design of circuits.

B. *Magnetic Regulation and Servo Applications*
Applications of magnetic amplifiers in the control of voltage, current, speed, temperature, frequency, position, etc.

C. *Magnetic Frequency Conversion*
Sub and ultraharmonic generators and magnetic frequency converters.

D. *Operational and Instrument Magnetic Amplifiers*
Metering and operational magnetic amplifiers and applications in analog computation. Magnetic modulators.

E. *Static Magnetic Switching Applications*
Digital computer applications and discrete step output magnetic amplifiers. Bistate magnetic amplifiers.

F. *Transistor-Magnetic Device Combinations*
All applications where transistors are used in conjunction with magnetic devices to provide new capabilities.

G. *Magnetism and Magnetic Materials*
Magnetic phenomena and the properties of magnetic amplifier materials.

H. *Measurements and Testing*
Techniques for the measurement of magnetic properties, core matching, and grading. Methods for measuring magnetic amplifier performance.

J. *Associated Devices*
Components used in conjunction with saturable reactors in magnetic amplifier circuitry. Devices generally associated with magnetic amplifier technique such as rectifiers, reference elements, etc.

K. *Ferroresonance and Its Applications*
The phenomenon of ferroresonance and its applications in regulation and harmonic frequency generation.

L. *Manufacturing Techniques, Environmental Effects, and Reliability*
Manufacturing and packaging techniques utilized to obtain a product reliable over a wide range of environmental conditions. The effects of radiation and temperature on magnetic amplifier performance.

Each subject heading is further subdivided into years. For each year, the papers under a given subject classification are listed alphabetically by author, except for patents which are listed in order of increasing patent numbers.

A list of abbreviations of the journals referred to in the bibliography and a list of general references on magnetic amplifiers follows.

Paper 58-1328, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted August 7, 1958; made available for printing September 11, 1958.

The members of the Bibliography Working Group of the Applications Subcommittee of the AIEE Magnetic Amplifiers Committee are E. J. ALEXANDER, *chairman;* V. B. Boros, W. A. Geyger, R. B. Morgan, and R. A. Raber.

List of Journals and Their Abbreviations

AEG Mitt. (Berlin)	*AEG Mitteilungen.* Hohenzollerndamm 150, Berlin-Grunewald, Germany
AIEE Trans.	*Transactions*, American Institute of Electrical Engineers, 33 W. 39th Street, New York 18, N. Y. (*Communication and Electronics* is Part I of the *Transactions*)
ASEA J.	*ASEA Journal.* ASEA Electric Ltd., Fulbourne Road, London, E. 17, England; Allmänna Svenska Elektriska AB, Vasteräs, Sweden
Automation and Remote Control (USSR)	*Avtomatika i Telemekhanika.* Institute of Automatics and Telemechanics, Mowcow Center, Mal. Kharitonievsky, Per 4, Moscow, USSR
BTH Activ.	*BTH Activities.* British Thomson-Houston Company, Ltd., Rugby, England
Bull. Soc. Franc. Élect	*Bulletin*, Société Francaise des Électriciens, 8–14 Avenue Pierre Larousse, Malakoff, Seine, France
C. R. Acad. Sci (Paris)	*Comptes Rendus*, Hebdomadaires des Séances, Académie des Sciences, Libraire Gauthier-Villars, Quai des Grands-Augustins 55, Paris (6ᵉ), France
Cont. Engg.	*Control Engineering.* McGraw-Hill Publishing Company, 330 W. 42nd Street, New York 36, N. Y.
Direct Current	*Direct Current.* Garraway Ltd., 11A Kensington Church Street, London, W. 8, England
Dtsch. Electrotechnik	*Deutsche Electrotechnik.* Verlag Technik, Unter den Linden 12, Berlin, N.W. 7, Germany
Elect. Energy	*Electrical Energy.* Morgan Brothers, 28 Essex Street, Strand, London, W.C. 2, England
Elect. Engg.	*Electrical Engineering.* American Institute of Electrical Engineers, 33 W. 39th Street, New York, N. Y.
Elect. Mfg.	*Electrical Manufacturing.* Gage Publishing Co., Inc., 1250 Sixth Avenue, New York 20, N. Y.
Electricity (USSR)	*Elektrichestvo.* Cherkasski Per, Moscow, USSR
Electronics	*Electronics.* McGraw-Hill Publishing Company, Inc., 330 W. 42nd Street, New York, N. Y.; Aldwych House, London, W.C. 2, England
Electrotechnika (Budapest)	*Electrotechnika.* V. Honved-u. 22, Budapest, Hungary
Elect. Times	*Electrical Times.* Sardinia House, Sardinia Street, London, W.C. 2, England
Elekt. Stantsii (USSR)	*Elektricheskie Stantsii.* Ministry of Electric Power Stations, GUM, III Korpus, Vetoshnyi Per, Moscow, USSR
EM (Vienna)	*Elektrotechnik und Maschinenbau.* Springer-Verlag, Molkerbastei 5, Vienna I, Austria
Energia Elett.	*Energia Elettrica.* Via Revere 14, Milan, Italy
Engineer	*Engineer.* 28 Essex Street, London, W.C. 2, England
ETZ	*Elektrotechnische Zeitschrift.* Verband Deutscher Elektrotechniker, Friedrich Ebert-Str. 111, Wuppertal-Elberfeld, Germany
Indust. Ital. Elettrotec.	*Industria Italiana Elettrotecnia.* Via Donizetti 30, Milan, Italy
Instrum. & Automation	Title extended from *Instruments*, 1954
Instruments	*Instruments.* Instruments Publishing Company, 921 Ridge Avenue, Pittsburgh 12, Pa.
IRE Conv. Record	*IRE Convention Record.* Institute of Radio Engineers, 1 E. 79th Street, New York, N. Y.
Iron & Steel Engr.	*Iron and Steel Engineer.* Association of Iron and Steel Engineers, 1010 Empire Building, Pittsburgh 22, Pa.
J. Brit. Instn. Radio Engrs.	*Journal*, British Institution of Radio Engineers, 9 Bedford Square, London, W.C. 1, England
J. Res. Nat. Bur. Stand.	*Journal of Research*, National Bureau of Standards, U. S. Government Printing Office, Washington 25, D. C.
Mullard Tech. Commun.	*Mullard Technical Communications.* Century House, Shaftesbury Avenue, London, W.C. 2, England
Nachrichtentechnik (Berlin)	*Nachrichtentechnik.* Verlag Technik, Unter den Linden 12, Berlin, N.W. 7, Germany
Nature	*Nature.* Macmillan & Company, Ltd., St. Martin's Street, London W.C. 2, England
OTS	*Office of Technical Services, Reports.* Department of Commerce, Washington 25, D. C.
Philips Tech. Rev.	*Philips Technical Review.* Philips' Gloeilampenfabrieken, Eindhoven, Holland
Proc. 1957 Elect. Comp. Symposium	*Proceedings*, 1957 Electronic Components Symposium, Institute of Radio Engineers, 1 E. 79th Street, New York 21, N. Y.
Proc. IEE	*Proceedings*, Institution of Electrical Engineers, Savoy Place, London, W.C. 2, England
Proc. IRE	*Proceedings*, Institute of Radio Engineers. 1 E. 79th Street, New York 21, N. Y.
Proc. STC on Mag. Amp., T-98	*Proceedings*, Special Technical Conference on Magnetic Amplifiers, *AIEE Special Publication T-98*, August 1957. American Institute of Electrical Engineers, 33 W. 39th Street, New York 18, N. Y.
Radio Engg. (USSR)	*Radiotekhnika.* ul. Kirova, 40, Moscow, USSR
Researches Electrotech. Lab. (Tokyo)	*Researches*, Electrotechnical Laboratory Agency of Industrial Science and Technology, Ministry of International Trade and Industry, 2-Chrome, Nagato-Machi, Chiyoda Ku, Tokyo, Japan
Rev. Élect. Mecan.	*Revue d'Électricité et de Mécanique.* Société Générale de Constructions Électriques et Mécaniques Alsthom, 38 Avenue Kléber, Paris (XVIᵉ), France
Siemens-Zs	*Siemens-Zeitschrift.* Siemens-Halske, Werner-Von Siemens-Str. 50, Erlangen, Germany
Trans., PG., IRE	*Transactions*, Professional Groups, Institute of Radio Engineers, 1 E. 79th Street, New York 21, N. Y.
VDE Fachberichte	*VDE Fachberichte.* VDE-Verlag, Friedrich-Ebert-Strasse 111, Wuppertal-Elberfeld, Germany
Westinghouse Engr.	*The Westinghouse Engineer.* 306 Fourth Avenue, Pittsburgh, Pa.

General References on Magnetic Amplifiers

1. Saturating Core Devices (book), L. R. Crow. Scientific Book Publishing Company, Vincennes, Ind., 1949.

2. Magnetic Amplifiers (book), G. M. Ettinger. Methuen, London, England; John Wiley & Sons, Inc., New York, N. Y., 1953, 88 pp.

3. Magnetic Amplifiers (book, in Czech), M. Gabler, J. Haskovec, E. Tomanek. Prague, Czechoslovakia, 1956, 251 pp.

4. Magnetic Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, 330 W. 42nd St., New York 36, N. Y., second edition, 1957, 394 pp.

5. The Transductor Amplifier (book), U. H. Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947, 178 pp.

6. The Transductor (book), A. U. Lamm. Esselte Aktiebolag, Stockholm, Sweden, second edition, 1948.

7. Transductors and Magnetic Amplifiers (book), A. G. Milnes. Macmillan & Company, Ltd., London, England, 1957, 286 pp.

8. The Magnetic Amplifier (book), J. H. Reyner. Stuart and Richards, London, England, 1950, 119 pp.

9. Magnetic Amplifiers and Saturable Reactors (book), M. G. Say, editor. George Newnes, London, England, 1954, 200 pp.

10. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, 545 pp.

11. Proceedings of the Symposium on Nonlinear Circuit Analysis, April 1956. Polytechnic Institute of Brooklyn; Interscience Publishers, New York, N. Y., 456 pp.

12. Bibliography on Transductors and Magnetic Amplifiers, H. B. Rex. *Instruments*, vol. 21, no. 4, 1948.

13. Bibliography of Magnetic Amplifiers Devices and Saturable Reactor Art, J. G. Miles. *AIEE Trans.*, vol. 70, pt. II, 1951, pp. 2104–23, 901 references.

14. Bibliography on Magnetic Amplifiers for 1951–1954. *Automation and Remote Control* (USSR), vol. 17, no. 5, May 1956, pp. 471–87.

15. Bibliography on Magnetic Amplifiers for 1955, G. V. Subbotina. *Ibid.*, no. 9, Sept. 1956, pp. 858–64.

16. Magnetic Amplifier Bibliography, 1951–1956, AIEE Committee Report. *AIEE Trans.*, vol. 77, pt. I, Nov. 1958, pp. 613–27.

I: English Language Articles

A. Theory, Analysis, and Design

1957

17. Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load, R. E. Anderson. *AIEE Trans.*, vol. 77, pt. I, Mar. 1957, pp. 76–86.

18. A Transient Analysis of a Bistable Magnetic Amplifier, A. Aronowitz. *Report no. R-592-57*, Microwave Research Institute, Polytechnic Institute of Brooklyn, N. Y., July 10, 1957, 44 pp.

19. The Series Magnetic Amplifier, Part I—The Four Modes of Operation, R. C. Barker. *AIEE Trans.*, vol. 76, pt. I, 1956 (Jan. 1957 section), pp. 819–25.

20. The Series Magnetic Amplifier, Part II—The Mechanisms of Gain and Time Constant, R. C. Barker. *Ibid.*, pp. 825–31.

21. Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions, H. C. Bourne, Jr., D. Nitzan. *Ibid.*, vol. 77, pt. I, Nov. 1957, pp. 646–55.

22. Dynamic Operation of Magnetic Amplifiers for Feedback Control Systems, H. C. Bourne, Jr., T. Kadota, D. Nitzan. *Scientific Report no. 4*, Electronics Research Laboratory, University of California, Berkeley, May 1957, 27 pp.

23. Elimination of Magnetic-Amplifier Control Circuit Inductance, D. P. Chandler, R. W. Downing. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 21; also *AIEE Trans.*, vol. 77, pt. I, Nov. 1958, pp. 550–57.

24. Some Aspects of Half-Wave Magnetic Amplifiers, G. M. Ettinger. *Proc. IEE*, Apr. 1957, 12 pp.

25. Introducing Young Engineers to the Appreciation of Magnetic Amplifier Problems, L. A. Finzi. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 51; also *AIEE Trans.*, vol. 77, pt. I, Mar. 1958, pp. 119–26.

26. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *AIEE Trans.*, vol. 77, pt. I, May 1957, pp. 229–40.

27. A General Circuit Theorem on Rectification, J. W. Gewartowski. *Proc. IRE*, vol. 45, Oct. 1957, p. 1410.

28. Generalized Transfer Characteristics for 3-Phase Magnetic Amplifiers with Inductive Loads, C. H. Grace. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 169.

29. A Half-Wave Bridge Magnetic Amplifier, W. Gregson. *Elect. Energy*, vol. 1, no. 7, Mar. 1957, pp. 200–06.

30. Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques, R. M. Hubbard. *AIEE Trans.*, vol. 77, pt. I, Nov. 1957, pp. 578–88.

31. A Comprehensive Study of Magnetic-Amplifier Design, Part I: A Practical Design Method. Part II: Toroidal-Core Dimensional Ratios as a Factor in Magnetic Amplifier Design, R. M. Hubbard, M. M. Bishop. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 121; also *AIEE Trans.*, vol. 77, pt. I, Nov. 1958, pp. 562–83.

32. Design of Minimum Weight Magnetic Cores, J. W. Kallander. *OTS, PB 121776*, Naval Research Laboratory, Jan. 1957, 8 pp.

33. On the Input Power of the Half-Wave Magnetic Amplifier Circuit, T. Kikuchi, K. Murakami. *AIEE Trans.*, vol. 76, pt. I, Mar. 1957, pp. 10–14.

34. Analysis of the Full-Wave Magnetic Amplifier Circuit Considering the Change of the Width of the Dynamic Hysteresis Loop, T. Kikuchi. *Ibid.*, July, pp. 241–48.

35. A Synchronous Power Converter for High-Efficiency Magnetic Amplifiers, J. D. Kopsho, Jr. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 230.

36. A-C Controlled Magnetic Amplifiers, E. W. Lehtonen, E. A. Cronauer. *Ibid.*, p. 8; also *AIEE Trans.*, vol. 77, pt. I, Sept. 1958, pp. 476–80.

37. Triggering in the Reversible Half-Wave Magnetic Amplifier, D. L. McMurtrie, J. P. Ward. *Ibid.*, p. 78; *AIEE Trans.*, Nov. 1958, pp. 598–605.

38. New Techniques in Designing Three-Phase Magnetic Amplifiers, W. H. Moore. *Ibid.*, p. 177.

39. Stabilized Magnetic Amplifier Circuits, H. W. Patton. *IRE Conv. Record*, pt. 6, 1957, p. 170.

40. Volt Second Transfer Efficiency in Fast Response Magnetic Amplifiers. Part 1: N^2/R and Control, T. J. Pula. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 107; also *AIEE Trans.*, vol. 77, pt. I, 1958 (Jan. 1959 section), pp. 861–67.

41. Current Steering in Magnetic Circuits, J. A. Rajchman, H. D. Crane. *Trans., PG., IRE*, Professional Group on Electronic Computers, vol. EC-6, no. 1, Mar. 1957, pp. 21–30.

42. The Simple Reactor Circuit; Its Operation and Mode Transition, J. F. Ringelman, A. L. Fenaroli. *AIEE Trans.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 660–68.

43. The Effect of Capacitance on Cross-Conduction in Half-Wave Bridge Magnetic Amplifiers, N. A. Ruggles. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 97.

44. Magnetic-Pulse-Generator Practical Design Limitations, M. F. Thompson, R. R. Trautwein, E. R. Ingersoll. *AIEE Trans.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 786–92.

45. Design Methods for Magnetic Amplifiers and Saturable Reactors, J. R. Walker, M. Frank. Wayne Engineering Research Institute, Detroit, Mich. May 1957, 60 pp. (Supplement no. 1 to *WADC Technical Report no. 56-281, AD-110 431*).

B. Magnetic Regulation and Servo Applications

1956

46. Reset Magnetic Servo Amplifiers, M. Weiss. *Cont. Engg.*, vol. 3, no. 10, Oct. 1956, pp. 88–91.

47. Speed Control of Induction Motors Using Saturable Reactors, P. L. Alger, Y. H. Ku. *AIEE Trans.*, vol. 75, pt. III, 1956 (Feb. 1957 section), pp. 1335–41.

48. Solid-State Amplifiers in Control, J. A. Baring. *Instrum. & Automation*, vol. 30, no. 7, July 1957, pp. 1288–91.

49. High-Efficiency Push-Pull Magnetic Amplifier and Its Use as a D-C Shunt Motor Drive, C. L. Boyajian. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 35; also *AIEE Trans.*, vol. 77, pt. I, Nov. 1958, pp. 545–50.

50. A Simple High-Performance Magnetic Amplifier Generator Voltage Regulator, W. E. Cronquist. *Ibid.*, p. 245.

51. Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers, G. W. Dick, P. P. Biringer, G. R. Slemon. *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 497–503.

52. A Magnetic Amplifier Voltage-Overload Protection Circuit, R. J. Distler, C. Sturzenbecker. *Ibid.*, pp. 602–06.

53. Magnetic Amplifiers for Tension Control, R. E. Hull. *Elect. Mfg.*, vol. 59, no. 4, Apr. 1957, pp. 152–55.

54. New DC Power Supply for Aircraft, R. E. King, V. Janonis. *Westinghouse Engr.*, vol. 17, no. 3, May 1957, pp. 82–85.

55. A Magnetic Pulse-Current Regulator, J. D. Lawrence, Jr., T. H. Bonn. *IRE Conv. Record*, pt. 4, 1957, p. 102.

56. Speed Control of a D-C Motor Using a Magnetic Amplifier, W. Leonhard. *AIEE Trans.*, vol. 76, pt. I, May 1957, pp. 112–19.

57. Six Control Problems Solved by Magnetic Amplifiers, A. E. Maine. *Cont. Engg.*, Dec. 1957, 7 pp.

58. The D-C Controlled A-C Voltage Source, A New Magnetic Amplifier, E. W. Manteuffel, R. O. McCary. *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 562–66.

59. A Constant Voltage DC Source, W. L. Marks. *Elect. Energy*, July 1957, 4 pp.

60. Magnetically Regulated Reference Power Supply, R. A. Raber. *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 534–37.

61. High-Frequency Magnetic Servo Amplifier for Air-Borne Control Systems, W. R. Seegmiller. *Ibid.*, pp. 526–29.

62. Applications of Non-linear Magnetics, Herbert F. Storm. *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 205; also *AIEE Trans.*, vol. 77, pt. I, July 1958, pp. 380–88.

63. Voltage Regulation, J. R. Walker. *Automation*, vol. 4, no. 4, Apr. 1957, pp. 73–74.

64. Transistor Magnetic Amplifier for Power Motor Control, S. Weissman, M. Bracutt. *Elect. Mfg.*, vol. 59, no. 6, June 1957, pp. 157–63.

C. Magnetic Frequency Conversion

1956

65. A Pulse Generator Using Nonlinear Magnetic Cores, W. B. Bridges. *Elect. Engg.*, vol. 75, no. 10, Oct. 1956, pp. 949–53.

66. Principle and Analysis of a Stabilized Phase Multiplier Type of Magnetic Fre-

I: English Language Articles

C. Magnetic Frequency Conversion

1956 (cont'd.)

quency Converter, **E. Friedlaender.** *Elect. Energy*, vol. 1, no. 2, Oct. 1956, pp. 55–60.

1957

67. Two Phase Motor Drive Using Static Frequency-Doubler Amplifiers, **G. W. Dick, P. P. Biringer, G. R. Slemon.** *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 497–503.

68. Static High Frequency Generator and Magnetic Amplifier, **M. Frank, J. R. Walker.** *OTS PB131240*, Wayne Engineering Institute, Feb. 1957, 202 pp.

69. Circuit for a Magnetic Subharmonic Pulser, **K. C. Hu, Y. H. Ku.** *AIEE Trans.*, vol. 76, pt. I, May 1957, pp. 137–40.

70. Magnetic Frequency Multipliers and Their Rating—Part II, Frequency Sextuplers, **W. McMurray.** *Ibid.*, July, pp. 289–93.

71. Magnetic-Pulse-Generator Practical Design Limitations, **M. F. Thompson, R. R. Trautwein, E. R. Ingersoll.** *Ibid.*, vol. 75, pt. I, 1956(Jan. 1957 section), pp. 786–92.

D. Operational and Instrument Magnetic Amplifiers

1956

72. USAF Computer Uses Magnetic Amplifiers. *Electronics*, vol. 29, no. 7, July 1956, pp. 138–39.

73. Push-Pull Magnetic Amplifier with DC Output, **R. Kh. Balyan.** *Direct Current*, vol. 3, no. 2, Sept. 1956, pp. 58–63.

1957

74. A Two-Megawatt Magnetic Modulator, **J. Antin.** *Report no. R-551-57; PIB-479; Technical Note no. RADC-TN-57-188*, Microwave Research Institute, Polytechnic Institute of Brooklyn, Feb. 27, 1957, 59 pp.

75. A Direct Current Transformer with Adjustable Ratio, **H. E. Darling.** *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 236.

76. Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type, **W. A. Geyger.** *AIEE Trans.*, vol. 76, pt. I, Mar. 1957. pp. 39–46.

77. Sixty-Cycle Low-Level Self-Balancing Magnetic Amplifier of the Differential-Feedback Type, **W. A. Geyger.** *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 187.

78. Operational Magnetic Amplifiers with Multiple Control Windings, **T. T. Kadota, H. C. Bourne.** *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 515–20.

79. The Operation of the Current-Type Self-Balancing Magnetic Amplifier, **A. D. Krall.** *Ibid.*, Sept., pp. 380–84.

80. Magnetically Keyed, Phase-Sensitive Demodulators, **R. B. Mark, W. X. Johnson, P. R. Johannessen.** *Ibid.*, Mar., pp. 1–5.

81. The Application of Miniature Saturable Reactors to Electronic Instruments, **R. S. Melsheimer.** *Trans., PG., IRE*, Professional Group on Instrumentation, vol. I-6, no. 2, June 1957. p. 148.

E. Static Magnetic Switching Applications

1956

82. A Magnetically Controlled Gating Element, **D. A. Buck.** Proceedings of the Eastern Joint Computer Conference, *AIEE Special Publication T-92*, Dec. 1956, pp. 47–50.

83. A Saturable-Transformer Digital Amplifier with Diode Switching, **E. W. Hogue.** *Ibid.*, pp. 58–64.

1957

84. Multihole Ferrite Core Configurations, **H. W. Abbott, J. J. Suran.** *Proc. IRE*, Aug. 1957, 13 pp.

85. New Static Control System, **F. C. Fennell.** *Iron & Steel Engr.*, vol. 34, no. 8, Aug. 1957, pp. 125–31.

86. Magnetic Output Amplifiers for Digital Control, **A. Kernick, H. A. Perkins, Jr.** *AIEE Trans.*, vol. 76, pt. I, Nov. 1957, pp. 504–08.

87. Principles of Transfluxor and Core Circuits, **J. A. Rajchman.** *Ind. Serv. Lab RB-120*, RCA Laboratories Division, Princeton, N. J., Sept. 10, 1957, 15 pp.

88. A Survey of Magnetic and Other Solid-State Devices for the Manipulation of Information, **J. A. Rajchman.** *Trans., PG., IRE*, Professional Group on Circuit Theory, vol. CT-4. no. 3, Sept. 1957, p. 210.

89. A Magnetic Amplifier Limit Switch, **D. L. Watrous.** *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 256.

90. Magnetic Amplifier Step Regulator, **B. Webb, Jr., G. D. Throop.** *Elect. Engg.*, vol. 76, no. 11, Nov. 1957, p. 957.

F. Transistor-Magnetic Device Combinations

1956

91. The Design and Operation of Transistor DC Converters. *Mullard Tech. Commun.*, vol. 2 Feb. 1956, pp. 161–204.

92. D. C. Transformers, **J. S. Smith.** *Trans., PG., IRE*, Professional Group on Vehicular Communication, vol. PGVC-6, July 1956, pp. 53–56.

1957

93. A Phase-Regulated Transistor Power Supply, **D. E. Deuitch, H. J. Paz.** *Trans., PG., IRE*, Professional Group on Circuit Theory, vol. CT-4, no. 3, Sept. 1957, pp. 279–84.

94. Transistor-Magnetic Amplifiers, **N. Jasper, J. C. Taylor, W. T. White.** *Elect. Mfg.*, vol. 60, no. 3, Sept. 1957, pp. 82–88.

95. Magnetic Regulation Transistor Power Supply, **L. F. Lyons.** *AIEE Trans.*, vol. 75, pt. I, 1956(Jan. 1957 section), pp. 643–45.

96. Development of the Transistor Inverter at 20 KC Using Power Transistors, **W. A. Martin.** *Trans., PG., IRE*, Professional Group on Instrumentation, vol. I-6, no. 2, June 1957, pp. 118–22.

97. A High-Speed Two-Winding Transistor-Magnetic-Core Oscillator, **A. J. Meyerhoff, R. M. Tillman.** *Ibid.*, Professional Group on Circuit Theory, vol. CT-4, no. 3, Sept. 1957, pp. 228–35.

98. Transistor Circuits and Applications, **A. G. Milnes.** *Proc. IEE*, vol. 104, pt. B, Nov. 1957, pp. 565–80.

99. The Balanced Transistor DC Converter, **J. Noordanus.** *Philips Tech. Rev.*, vol. 18, no. 3, Sept. 1957, pp. 125–36.

100. Design Considerations of Junction-Transistor Oscillators for the Conversion of Power from Direct to Alternating Current, **F. Oakes.** *Proc. IEE*, vol. 104, pt. B, May 1957, pp. 307–17.

101. A New Family of Transistor Switching Circuits, **M. Rubinoff.** *AIEE Trans.*, vol. 76, pt. I, July 1957, pp. 286–89.

102. Transistor Magnetic Amplifier for Power Motor Control, **S. Weissman, M. Bracutt.** *Elect. Mfg.*, vol. 59, no. 6, June 1957, pp. 157–63.

G. Magnetism and Magnetic Materials

1957

103. Magnetic Alloy of High Permeability and Low Hysteresis Loss. *Engineer*, vol. 203, Mar. 15, 1957, p. 427.

104. Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores, **AIEE Committee Report.** *AIEE Trans.*, vol. 76, pt. I, May 1957, pp. 135–37.

105. Dynamic Core Behavior and Magnetic Amplifier Performance, **L. A. Finzi, D. L. Critchlow.** *AIEE Trans.*, vol. 76, pt. I, May 1957, pp. 229–40.

106. Supermendur, A New Rectangular-Loop Magnetic Material, **H. L. B. Gould, D. H. Wenny.** *Elect. Engg.*, vol. 76, Mar. 1957, pp. 208–11.

107. Sendust Flake—A New Magnetic Material for Low Frequency Application, **W. M. Hubbard, E. Adams, J. F. Haben.** *Trans., PG., IRE*, Professional Group on Component Parts, vol. CP-4, no. 1, Mar. 1957, p. 2.

108. Design of Minimum Weight Magnetic Cores, **J. W. Kallander.** *OTS, PB121776*, Jan. 1957, [illegible] pp.

109. Survey of Square Loop Magnetic Materials, **V. E. Legg.** *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 69.

110. Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores, **P. L. Schmidt.** *Elect. Mfg.*, vol. 60, no. 3, Sept. 1957, p. 122–24.

111. Applying High Permeability Laminations, **W. S. Spring.** *Elect. Mfg.*, vol. 59, no. 3, Mar. 1957, pp. 126–31.

H. Measurements and Testing

1956

112. Investigation of an Alternating Current Bridge for the Measurement of Core Losses in Ferromagnetic Materials at High Flux Densities, **I. L. Cooter, W. P. Harris.** *J. Res. Nat. Bur. Stand.*, vol. 57, no. 2, Aug. 1956, pp. 103–12.

1957

113. Analysis of 3-Leg Transformer Cores and a New Instrument for Measurement of Exciting-Current Waveforms, **J. K. Delson.** *AIEE Trans.*, vol. 76, pt. III, Dec. 1957, pp. 1104–12.

114. Semiautomatic Test for Measuring Loss in Transformer Core Laminations, **S. C. Leonard, R. L. Tenbroeck.** *Ibid.*, Apr. 1957, pp. 118–24.

115. A Power Method of Measuring the Torque-Producing Output of a Magnetic Amplifier Driving One Phase of a Two-Phase Servo Motor, **S. J. Reisman, E. H. Bowman.** *Proc. STC on Mag. Amp., T-98*, Aug. 1957, p. 45; also *AIEE Trans.*, vol. 77, pt. I, Nov. 1958, pp. 584–86.

116. Core Tester Simplifiers Ferro-Amplifier Design, **R. G. Roberts, C. C. Horstman.** *Electronics*, vol. 30, no. 8, Aug. 1, 1957, pp. 150–53.

J. Associated Devices

1957

117. A Ceramic–Metal Voltage Reference Tube, **J. W. Culp, P. Koskos.** *Trans., PG., IRE*, Professional Group on Electron Devices, vol. ED-4, no. 2, Apr. 1957, p. 144.

118. Characteristics of Silicon Junction Diodes as Precision Voltage Reference Devices, **K. Enslein.** *Ibid.*, Professional Group on Instrumentation, vol. I-6, no. 2, June 1957, p. 105.

119. Rating and Application of Germanium and Silicon Rectifiers, **F. W. Gutzwiller.** *AIEE Trans.*, vol. 75, pt. I, 1956(Jan. 1957 section), pp. 753–57.

120. The Fused Silicon Rectifier, **H. W. Henkels.** *Ibid.*, pp. 733–45.

121. Silicon Rectifiers in Brushless Machinery, **H. W. Henkels, H. R. Camp, J. R. Fortier, W. J. Shilling, R. E. Smith.** *Ibid.*, vol. 76, pt. I, Nov. 1957, pp. 620–26.

122. Static D-C References for Closed-Loop Controls, **M. Mamon.** *Elect. Mfg.*, vol. 59, no. 1, Jan. 1957, pp. 54–61.

123. Properties of Ferroelectric Devices as Current Regulating and Frequency Determining Devices, **C. Rosenberg.** *Proc. 1957 Elect. Comp. Symposium*, p. 38.

124. Calculation of Inductance of Toroids with Rectangular Cross Section and Few Turns, **R. F. Schwartz.** *Proc. IRE*, vol. 45, Oct. 1957, pp. 1416–17.

125. Design Relationships for Iron-Core Filter Chokes, **T. Usher, Jr.** *AIEE Trans.*, vol. 76, pt. I, Sept. 1957, pp. 484–87.

126. Circuit Components for High Voltage DC Power Supplies, **V. Wouk.** *IRE Wescon Convention Record*, New York, N. Y., pt. 6, 1957, p. 30.

K. Ferroresonance and Its Applications

1955

127. Limiting Conditions in a Ferroresonant Trigger Circuit, **J. C. Akalnik.** *OTS, PB 124754*, Oct. 1955, 54 pp.

1956

128. Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis, **Egon Brenner.** *AIEE Trans.*, vol. 75, pt. I, Sept. 1956, pp. 450–56.

1957

129. A New Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits—Part I, Subharmonics do not Exist, **M. Kamal Gohar.** *AIEE Trans.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 771–75.

130. Ferroresonant Circuits for Digital Computers, **C. B. Newport, D. A. Bell.** *J. Brit. Instn. Radio Engrs.*, vol. 17, no. 11, Nov. 1957, pp. 619–30.

131. Discontinuous Transition Time Between Stable States in Ferroresonant Circuits, **S. Plotkin.** *AIEE Trans.*, vol. 76, pt. I, Sept. 1957, pp. 410–20.

132. Ferro-Resonance in Voltage-Transformers, **S. Rusck.** *ASEA J.*, vol. 30, no. 1–2, 1957, pp. 20–22.

L. Manufacturing Techniques, Environmental Effects, and Reliability

1957

133. Modern Methods of Winding Toroidal Coils. *Demodulator*, Feb. 1957, 5 pp.

134. Temperature Characteristics of the Transfluxor, **H. W. Abbott, J. J. Suran.** *Trans., PG., IRE*, Professional Group on Electron Devices, vol. ED-4, no. 2, Apr. 1957, p. 113.

135. Electronic Material at Elevated Temperatures, **J. W. Ballard.** *OTS PB 131157*, May 1957, 10 pp.

136. High Temperature Magnetic Amplifier, **M. Frank, J. R. Walker.** *WADC Tech. Report No. 57-74*, Wright Air Development Center, Dayton, Ohio, June 1957, 98 pp.

137. High Temperature Magnetic Amplifiers, **M. Frank, J. R. Walker.** *Proc. 1957 Elect. Comp. Symposium*, pp. 186–93.

138. Effects of Irradiation upon Diodes of the Silicon Junction Type, **R. Gorton.** *Nature*, vol. 179, Apr. 27, 1957, p. 864.

139. Ultra High Temperature (500°C) Miniaturized Power Transformers and Inductors, **H. B. Harms, W. R. Langdon, J. F. Rippin, G. E. Walter.** *Proc. 1957 Elec. Comp. Symposium*, pp. 177–85.

140. Effect of Nuclear Radiation on Selected Semiconductor Devices, **G. L. Keister, H. V. Stewart.** *Proc. IRE*, vol. 45, no. 7, July 1957, pp. 931–37.

141. Liquid and Gas Dielectrics for Magnetic Component Miniaturization, **L. F. Kilham, Jr., R. R. Ursch.** *Elect. Mfg.*, vol. 59, no. 5, May 1957, pp. 140–45.

142. An Analytical Investigation of the Effects of Nuclear Radiation on Transistors, **D. B. Kret.** *Proc. 1957 Elec. Comp. Symposium*, pp. 75–81.

143. Effects of Nuclear Radiation on Electronic Components and Systems, **J. R. Milliron.** *OTS, PB 131051*, Mar. 1957, 9 pp.

144. Designing Electronics to Resist Nuclear Energy, **H. L. Morgan.** *Electronics*, vol. 30, no. 5, May 1, 1957, pp. 155–57.

145. Organosilicon Compounds as Insulation for Miniature Power Transformers for High Temperature Operation, **F. T. Parr.** *Proc. 1957 Elec. Comp. Symposium*, pp. 246–53.

146. Effect of Radiation on Semiconductors, **J. C. Pigg, C. C. Robinson.** *Elect. Mfg.*, vol. 59, no. 4, Apr. 1957, pp. 116–24.

147. An Experimental Investigation of the Effects of Nuclear Radiation on Transistors, **A. J. Schwartz.** *Proc. 1957 Elec. Comp. Symposium*, pp. 67–74.

148. Effects of Nuclear Radiation on Electronic Components, **R. D. Shelton.** *Elect. Mfg.*, vol. 60, no. 3, Sept. 1957, pp. 76–81.

149. High-Temperature Magnetic Amplifiers and Transformers, **J. R. Walker, M. Frank.** *Ibid.*, vol. 59, no. 6, June 1957, pp. 168–73.

II: Foreign Language Articles

A. Theory, Analysis, and Design

1956

150. Seminar and Magnetic Amplifiers and Contactless Magnetic Elements. *Automation and Remote Control* (USSR), vol. 17, no. 12, Dec. 1956, p. 1275.

151. Calculation of Magnetic Amplifiers with Feedback, **R. Kh. Bal'yan.** *Electricity* (USSR), no. 9, 1956, pp. 63–67.

152. The Self-Saturating Magnetic Amplifier, **M. B. Chague.** *Bull. Soc. Franc. Elect.*, vol. 6, no. 61, Jan. 1956, pp. 14–56.

153. Low Power Magnetic Amplifiers, **R. Klein.** *Rev. Elect. Mecan.*, no. 106, July-Sept. 1956, pp. 1–28.

154. Input Resistance of Magnetic Amplifiers and Its Effect on Operational Properties, **F. Kümmel.** *VDE Fachberichte*, vol. 19 (I), 1956, pp. 152–59.

155. Push Pull Excitation of Electric Machines Using Magnetic Amplifiers, **F. Kümmel.** *ETZ*, vol. 77, no. 14, July 1956, pp. 492–98.

156. Graphical Analytical Method of Calculating the Characteristics of Push-Pull Magnetic Amplifiers, **M. E. Poyurovskii.** *Electricity* (USSR), no. 7, 1956, pp. 57–60.

157. Operation of a Saturable Reactor with a Semiconductor-Type Rectifier and an Active-Reactive Load, **A. G. Zdrok, G. P. Smirnov.** *Ibid.*, no. 10, pp. 44–47.

1957

158. Decreasing the Lag of Magnetic Amplifiers by Introducing Derivative Feedback, **O. I. Aven.** *Automation and Remote Control* (USSR), Feb. 1957, 8 pp.

159. The Design of an AC Magnetic Amplifier for Given Parameters, **B. K. Bool.** *Ibid.*, Oct. 1957, 4 pp.

160. Magnetic Amplifiers of Classical Design (in Hungarian), **F. Csaki, A. Frigyes.** *Elektrotechnika* (Budapest), vol. 50, no. 8–9, Aug.-Sept., 1957, pp. 294–308.

161. Methods of Modifying the Dynamic Response of Magnetic Amplifiers, **C. Delgrosso, G. Sacco.** *Elettrotecnica* (Milan), vol. 44, no. 6, June 1957, pp. 318–30.

162. A Three-Phase Magnetic Amplifier, **V. V. Gorsky.** *Automation and Remote Control* (USSR), July 1957, 5 pp.

163. Magnetic Amplifier Circuits with Proportional and Derivative Feedback, **L. A. Grigorian.** *Ibid.*, no. 6, June 1957, 14 pp.

164. A Design of an AC Magnetic Amplifier, **N. A. Kaluzhnikov.** *Ibid.*, Sept. 1957, 5 pp.

165. Simplified Design of Magnetic Amplifiers with Nickel-Iron Cores, **N. A. Kaluzhnikov.** *Ibid.*, Mar. 1957, 5 pp.

166. Principles of Magnetic Amplifier Engineering, **F. Kümmel.** *Nachrichtentechnik* (Berlin), vol. 10, no. 4, Apr. 1957, pp. 153–59.

167. Theory of a Half-Wave Magnetic Amplifier, **R. A. Lipman, I. B. Negnevitsky.** *Automation and Remote Control* (USSR), vol. 18, no. 5, Apr. 1957, pp. 449–65.

168. Certain Problems Associated with the Theory of Magnetic Amplifiers and Magnetically Modulated Probes of the Second Harmonic Type, **V. N. Mikhailovsky, Iu. I. Spektor.** *Ibid.*, Aug. 1957, 8 pp.

169. The Problem of Matching Second-Harmonic Magnetic Amplifiers to a Load, **V. N. Mikhailovsky, Iu. I. Spektor.** *Ibid.*, June 1957, 9 pp.

170. Characteristics and an Application of Saturable Reactors Using Silicon-Steel Cores, **K. Morisawa.** *Researches Electrotech. Lab.* (Tokyo), no. 557, Feb. 1957, 153 pp.

171. Frequency Characteristics of Magnetic Materials for Cores of Magnetic Amplifiers, **G. V. Subbotina.** *Automation and Remote Control* (USSR), Oct. 1957, 3 pp.

172. Special Considerations in Designing Multistage Magnetic Amplifiers, **N. P. Vasil'eva, M. A. Boiarchenkov.** *Ibid.*, July 1957, 9 pp.

173. Design of Magnetic Amplifiers for a Given Supply Voltage, **N. P. Vasil'eva, O. A. Sedykh.** *Ibid.*, vol. 18, no. 11, 1957, pp. 1052–60.

B. Magnetic Regulation and Servo Applications

1956

174. Magnetic Servo Amplifier with Minimum Response Time, **J. Auricoste.** *Automation*, no. 4, Apr. 1956, pp. 147–48.

175. Electromagnetic AC Voltage Stabilizer, **G. Beszedics.** *EM* (Vienna), vol. 73, no. 17, Sept. 1956, pp. 415–18.

176. Electromagnetic Voltage Regulators and Their Adjustment, **D. E. Buryakorskil.** *Elekt. Stantsii* (USSR), no. 10, 1956, pp. 26–31.

177. A New Type of AC Voltage Stabilizer with Saturated Magnetic Circuits, **L. Cibrario.** *Indust. Ital. Elettrotec.*, vol. 9, no. 9, Sept. 1956, pp. 319–21.

178. Voltage Regulation of Large Three-Phase Generators with Magnetic Amplifiers and Amplifier Machines, **H. Junior.** *VDE Fachberichte*, vol. 19 (I), 1956, pp. 160–66.

179. The Power Magnestat and Its Application in Dynamic Braking, **S. Rudzinski.** *BTH Activ.*, vol. 27, no. 5, Sept.-Oct. 1956, pp. 154–58.

180. Magnetic Amplifiers for the Control of Stage Lighting Installations, **H. Watzinger.** *ETZ*, vol. 8, no. 11, Nov. 21, 1956, pp. 407–11.

1957

181. Speed Regulation with Magnetic Amplifiers and the Relay 2-Position Regulator for Machine Tools, **R. Baisch.** *ETZ*, vol. 78, no. 18, Sept. 1957, pp. 646–49.

182. Speed Regulation of a DC Motor by Transductors and Semi-conductors, **K. N. Borisov.** *Electricity* (USSR), no. 2, 1957, pp. 40–43.

183. An Analytical Expression for the Static Characteristic of the Detector Unit of a Saturable-Reactor Controller, **E. Ia.**

II: Foreign Language Articles

B. Magnetic Regulation and Servo Applications

1957 (cont'd.)

Iakubaitis. *Automation and Remote Control* (USSR), Mar. 1957, 6 pp.

184. The Asynchronous Motor Controlled by Magnetic Amplifiers, **W. Kafka, W. Döll.** *ETZ*, vol. 78, no. 24, Dec. 1957, pp. 884–90.

185. The Dynamic Performance of Magnetic Control Apparatus for Grid-Controlled Gas-Discharge Tubes, **G. Kessler, K. W. Pflug.** *Dtsch. Electrotechnik*, vol. 11, no. 11, Nov. 1957, pp. 501–07.

186. Current Limiting by the Use of Transductors, **P. Kratz, A. Lang.** *AEG Mitt.* (Berlin), vol. 47, no. 5–6, May-June, 1957, pp. 140–44.

187. Speed Regulation of Three-Phase Induction Motors, Using a Bridge-Type Transducer, **V. S. Kulebakin, S. M. Domanitski.** *Automation and Remote Control* (USSR), no. 2, 1957.

188. Use of Magnetic Amplifier for Braking of 3-Phase Asynchronous Motors, **H. G. Lott.** *AEG Mitt.* (Berlin), vol. 47, no. 1–2, Jan.-Feb. 1957, pp. 19–23.

189. Synchronous Single-Machine Electric Drive with Magnetic Amplifier for Facsimile Apparatus, **O. B. Pevzner, K. A. Brusilovskiy.** *Elektrosvias* (USSR), no. 8, 1957, pp. 60–68.

190. Induction Motor Drives of Centrifugal Pumps with Saturated Reactor Control, **M. M. Sokolov, V. M. Terekhov.** *Electricity* (USSR), no. 9, 1957, pp. 14–16.

C. Magnetic Frequency Conversion

1957

191. Static Frequency Tripler and Its Industrial Use, **K. Kurutz.** *Electrotechnika* (Budapest), vol. 50, no. 3, 1957, pp. 99–105.

192. Magnetic Frequency Divider, **K. G. Mityushkin.** *Electricity* (USSR), no. 8, 1957, pp. 53–58.

D. Operational and Instrument Magnetic Amplifiers

1956

193. The Theory and Design of a Magnetic Modulator Operating on the Principle of Frequency Doubling (in Russian), **M. A. Rozenblat.** *Radio Engg.* (USSR), vol. 11, no. 8, 1956, pp. 36–51.

1957

194. Matching of a Second Harmonic Magnetic Amplifier Under Load, **V. M. Mikhailovskii, Iu. I. Spektor.** *Automation and Remote Control* (USSR), vol. 18, no. 6, 1957, pp. 551–59.

195. Certain Problems in the Theory of Magnetic Amplifiers and Magneto-Modulation Probes of the Second Harmonic Type, **V. M. Mikhailovskii, Iu. I. Spektor.** *Ibid.*, no. 8, 1957, pp. 716–23.

F. Transistor-Magnetic Device Combinations

1957

196. Semiconductor DC Converter, **G. S. Tsykin.** *Radio Engg.* (USSR), vol. 12, no. 12, 1957, pp. 56–62.

G. Magnetism and Magnetic Materials

1956

197. Progress in Magnetic Materials and Their Application, **F. Pawlek, K. Reichel.** *AEG Mitt.* (Berlin), vol. 46, no. 11–12, Nov.-Dec. 1956, pp. 337–48.

1957

198. Observations on the Behavior of Ferromagnetic Materials in Toroidal Rings, **G. Biorci, D. Prescetti.** *Energia Elett.*, vol. 34, no. 8, Aug. 1957, pp. 795–99.

199. New Design of Quality Transductor Cores, **U. Krabbe, G. H. Giesenhagen.** *ETZ* (Berlin), vol. 78, no. 19, Oct. 1, 1957, pp. 712–16.

200. Fundamental Properties of Permalloy 45 N of Soviet Make, **R. Pac.** *Prace ITR*, no. 3, 1957, 15 pp.

201. Magnetic After Effect in Hot-Rolled Silicon Steel Sheets, **A. Smolinski, M. Zbikowski.** *Ibid.*

H. Measurements and Testing

1956

202. A Whole Sheet Test Apparatus for Grading Transformer Sheets, **A. Fiebiger.** *Siemens Zs.*, vol. 30, no. 9, 1956.

K. Ferroresonance and Its Applications

1956

203. A New Type of AC Voltage Stabilizer with Saturated Magnetic Circuits, **L. Cibrario.** *Indust. Ital. Elettrotec.*, vol. 9, no. 9, Sept. 1956 pp. 819–21.

1957

204. Contactless Switching Devices Using Transformer Ferroresonant Circuits, **Ia. G. Koblents, D. A. Iakovenko.** *Elektrosvias* (USSR), no. 3, 1957, 10 pp.

205. On the Appearance of Abnormal States in Certain Ferro-Resonant Circuits, **M. Panet.** *C.R. Acad. Sci.* (Paris), vol. 245, no. 8, Aug. 1957, pp. 834–37.

206. The Condition of Maximum Accuracy for Ferroresonant Voltage Stabilizers, **Z. Zilberman.** *Radio Engg.* (USSR), vol. 12, no. 2, 1957, pp. 71–72.

III: Domestic Patents

A. Theory, Analysis, and Design

July 1956 to December 1956

207. High Gain Self-Saturating Magnetic Amplifier, **S. P. Jackson.** *No. 2,758,161*, Dec. 5, 1952–Aug. 7, 1956.

208. Magnetic Amplifier, **M. G. Tekosky.** *No. 2,758,162*, Mar. 2, 1951–Aug. 7, 1956.

209. Magnetic Amplifier, **M. Sakamoto.** *No. 2,760,148*, Mar. 26, 1953–Aug. 21, 1956.

210. Self-Excited Transductors, **L. Borg.** *No. 2,762,966*, June 26, 1951–Sept. 11, 1956.

211. Magnetic Amplifier, **C. J. Jensen.** *No. 2,763,828*, May 27, 1953–Sept. 18, 1956.

212. Reversible-Polarity Direct-Current Output Magnetic Amplifier Requiring Only One Alternating-Current Source, **K. L. Sanders, Jr.** *No. 2,764,726*, Nov. 22, 1952–Sept. 25, 1956.

213. Magnetic Amplifier, **V. J. Louden.** *No. 2,765,374*, Aug. 18, 1952–Oct. 2, 1956.

214. Power Transmission, **W. J. Dornhoefer.** *No. 2,765,436*, July 28, 1950–Oct. 2, 1956.

215. Ring Modulator Magnetic Amplifier, **H. H. Woerdemann.** *No. 2,767,257*, Apr. 10, 1952–Oct. 16, 1956

216. Multistage Magnetic Amplifier System, **L. H. Rowley, W R. Rauth, Jr.** *No. 2,767,372*, Oct. 14, 1952–Oct. 16, 1956.

217. D.C. Converter Circuits, **E. Obermaier.** *No. 2,769,138*, Apr. 25, 1951–Oct. 30, 1956.

218. Push-Pull Magnetic Amplifier, **W. R Seegmiller.** *No. 2,770,769*, Nov. 12, 1952–Nov. 13, 1956.

219. Vacuum-Tube Controlled Magnetic Amplifier, **C. W. Lufcy.** *No. 2,770,770*, May 22, 1953–Nov. 13, 1956.

220. Magnetic Amplifier, **M. Sakamoto.** *No. 2,773,131*, Apr. 6, 1953–Dec. 4, 1956.

221. Magnetic Amplifiers, **W. J. Dunnet.** *No. 2,773,133*, May 25, 1954–Dec. 4, 1956.

222. Magnetic Amplifiers, **W. J. Dunnet.** *No. 2,773,134*, May 25, 1954–Dec. 4, 1956.

223. Magnetic Amplifiers, **F. S. Malick.** *No. 2,773,235*, Sept. 2, 1953–Dec. 4, 1956.

1957

224. Parallel-Connected Magnetic Amplifier, **E. L. Walker.** *No. 2,777,021*, June 4, 1954–Jan. 8, 1957.

225. Magnetic Amplifier, **J. A. Fingerett, F. A. Hill.** *No. 2,777,073*, May 24, 1954–Jan. 8, 1957.

226. Multi Branch Magnetic Amplifier with Beneficially Interrelated Magnetic Path, **W. F. Schmidt, Jr.** *No. 2,778,987*, Apr. 19, 1954–Jan. 22, 1957.

227. Three-Phase Magnetic Amplifier, **W. R. Rauth, Jr.** *No. 2,779,910*, Sept. 18, 1952–Jan. 29, 1957.

228. Means for Reducing or Nullifying the Minimum Output Current of Magnetic Amplifiers, **P. M. Fischer.** *No. 2,779,911*, Feb. 12, 1954–Jan. 29, 1957.

229. Self-Saturating Reactor Circuit, **B. Lee.** *No. 2,780,770*, Apr. 21, 1953–Feb. 5, 1957.

230. Self-Saturating Reactor Circuit, **B. Lee.** *No. 2,780,772*, Apr. 21, 1953–Feb. 5, 1957.

231. Magnetic Amplifier Circuits, **D. W. Grant.** *No. 2,782,269*, June 28, 1955–Feb. 19, 1957.

232. Magnetic Amplifier Control Circuit **R. A. Ramey, Jr.** *No. 2,783,315*, July 20, 1951–Feb. 26, 1957.

233. Power Transmission, **C. Bennett, L. W. Buechler.** *No. 2,790,135*, Dec. 8, 1955–Apr. 23, 1957.

234. Magnetic Amplifier, **J. C. Karlson.** *No. 2,791,742*, May 21, 1953–May 7, 1957.

235. Magnetic Power Amplifier, **H. T. Mortimer.** *No. 2,799,738*, July 27, 1956–July 16, 1957.

236. Magnetic Amplifier, **J. N. Van Scoyoc.** *No. 2,794,165*, Aug. 8, 1952–May 28, 1957.

237. Magnetic Controller, **C. B. House.** *No. 2,795,753*, June 1, 1953–June 11, 1957.

238. Magnetic Amplifiers, **F. S. Malick, F. G. Timmel.** *No. 2,795,652*, Sept. 15, 1953–June 11, 1957.

239. Arrangement in Parallel Operation of Transductors, **G. Sonesson.** *No. 2,795,754*, July 30, 1953–June 11, 1957.

240. Magnetic Amplifier Demodulator, **A. A. Brown.** *No. 2,797,384*, Sept. 30, 1954–June 25, 1957.

241. Series-Transductor Apparatus, **M. Belamin.** *No. 2,800,623*, Sept. 25, 1952–July 23, 1957.

242. Magnetic Amplifier Circuit, **A. L. Bastian.** *No. 2,800,626*, Nov. 14, 1952–July 23, 1957.

243. Magnetic Amplifier Control Apparatus, **F. S. Malick.** *No. 2,802,169*, Nov. 5, 1954–Aug. 6, 1957.

244. Saturable Control System, **G. H. Dewitz.** *No. 2,802,185*, July 24, 1952–Aug. 6, 1957.

245. Saturable Core Apparatus, **G. H. Dewitz.** *No. 2,802,186*, Apr. 19, 1952–Aug. 6, 1957.

246. Power Transmission, **S. Steinitz.** *No. 2,807,753*, Nov. 23, 1954–Sept. 24, 1957.

247. Power Transmission, **S. Steinitz.** *No. 2,807,754*, Nov. 23, 1954–Sept. 24, 1957.

248. Self-Saturating Reactor Circuit, **W. F. Schmidt, Jr.** *No. 2,807,775*, July 28, 1952–Sept. 24, 1957.

249. Magnetic Amplifier System, **L. W.**

Buechler, W. F. Schmidt, Jr. *No. 2,807,776*, July 28, 1952–Sept. 24, 1957.

250. Two-Stage Magnetic Amplifier, S. Weissman. *No. 2,809,241*, July 12, 1955–Oct. 8, 1957.

251. Magnetic Amplifier, M. C. Creusere. *No. 2,810,519*, Aug. 13, 1954–Oct. 22, 1957.

252. Magnetic Amplifier, G. H. Dewitz. *No. 2,812,389*, Apr. 19, 1952–Nov. 5, 1957.

253. Voltage-Controlled Magnetic Amplifier, P. Siskind. *No. 2,813,246*, Apr. 19, 1955–Nov. 12, 1957.

254. Magnetic Modulator, B. L. Freeman, B. H. Vester, Jr. *No. 2,814,738*, Feb. 13, 1957–Nov. 26, 1957.

255. Magnetic Amplifier Circuit Using Complex Feedback, E. V. Weir. *No. 2,817,807*, Aug. 11, 1950–Dec. 24, 1957.

B. Magnetic Regulation and Servo Applications

July 1956 to December 1956

256. Minimum Excitation Limit Circuits, J. T. Carleton, D. A. Burt. *No. 2,757,332*, Dec. 24 1953–July 31, 1956.

257. Control Circuit, F. J. Rau, K. Losch. *No. 2,760,083*, Dec. 8, 1954–Apr. 21, 1956.

258. Magnetic Limiter, M. J. Stateman. *No. 2,762,967*, Mar. 2, 1953–Sept. 11, 1956.

259. Output Mean Square Sensing for Magnetic Output Voltage Regulator, J. V. Majercak. *No. 2,762,968*, Sept. 20, 1955–Sept. 11, 1956.

260. Magnetic Phase and Amplitude Controller, J. A. Fingerett, J. E. Richardson. *No. 2,762,969*, Mar. 23, 1954–Sept. 11, 1956.

261. Servo System with Magnetic Amplifier with Integral Feedback, H. H. Woodson. *No. 2,764,719*, July 15, 1953–Sept. 25, 1956.

262. Magnetic Servo Amplifier, D. G. Scorgie. *No. 2,764,723*, Dec. 1, 1953–Sept. 25, 1956.

263. Saturable Core Reactor with Thermistor Control, P. R. Marvin. *No. 2,765,119*, Mar. 7, 1951–Oct. 2, 1956.

264. Saturable Reactor Control, T. F. McHenry. *No. 2,765,438*, Jan. 8, 1953–Oct. 2, 1956.

265. Voltage Regulators, Y. A. Rocard. *No. 2,765,439*, Oct. 29, 1949–Oct. 2, 1956.

266. Magnetic Amplifier Motor Control System, C. A. Schurr. *No. 2,766,415*, Mar. 12, 1953–Oct. 9, 1956.

267. Generator Frequency Control by Electric Braking, L. Black. *No. 2,767,367*, Mar. 3, 1953–Oct. 16, 1956.

268. Regulating System for Generators, P. W. Franklin. *No. 2,769,133*, Jan. 2, 1951–Oct. 30, 1956.

269. Magnetic Amplifier Control Device for an Electromagnetic Clutch, G. W. Allen. *No. 2,769,931*, Mar. 15, 1955–Nov. 6, 1956.

270. Monitor Control System, F. S. Malick. *No. 2,774,022*, June 15, 1953–Dec. 11, 1956.

271. Regulated Rectifying Apparatus, W. H. Bixby. *No. 2,774,930*, July 18, 1955–Dec. 18, 1956.

1957

272. Saturable Reactor Ward Leonard Control System, D. W. Fath, F. S. Horner, Jr. *No. 2,783,424*, June 11, 1954–Feb. 26, 1957.

273. Voltage Regulator, A. G. McKeena, K. H. Beeber. *No. 2,791,740*, Oct. 20, 1953–May 7, 1957.

274. Reversing Magnetic Amplifier Control System, G. R. Markow. *No. 2,792,541*, Mar. 23, 1953–May 14, 1957.

275. Magnetic Amplifier for Control Purposes, W. F. Horton, S. J. Reisman, R. O. Decker, R. A. Ramey. *No. 2,792,547*, Nov. 12, 1954–May 14, 1957.

276. Means for Eliminating Asymmetry Zero-Drift Errors in Magnetic Servo Amplifiers, W. A. Geyger. *No. 2,793,336*, Mar. 28, 1955–May 21, 1957.

277. Precision Frequency Regulator, E. C. Rhyne, Jr. *No. 2,793,342*, Dec. 22, 1955–May 21, 1957.

278. System for Controlling Induction Motors by Saturable Reactors and Coordinately Controlled Resistors, E. C. Rhyne, Jr. *No. 2,793,338*, Dec. 22, 1955–May 21, 1957.

279. Amplifier Apparatus for Motor Control, J. P. Montgomery, D. A. Poepsel, W. M. Brittain. *No. 2,795,751*, Apr. 14, 1955–June 11, 1957.

280. Magnetic Frequency and Voltage Control for Motor Generator, H. H. Woerdemann. *No. 2,797,382*, May 2, 1952–June 25, 1957.

281. Magnetic Amplifiers and Control Systems, D. H. Carlisle. *No. 2,798,170*, July 7, 1954–July 2, 1957.

282. Battery Charger and Circuit Therefor, M. Mamon. *No. 2,802,165*, June 11, 1954–Aug. 6, 1957.

283. Regulators, J. C. Sanderlin, L. F. Deise. *No. 2,802,166*, Feb. 20, 1956–Aug. 6, 1957.

284. Position Servo, J. E. Brook. *No. 2,802,976*, Nov. 1, 1954–Aug. 13, 1957.

285. Regulator for an Alternator, R. P. Evans. *No. 2,806,990*, Aug. 21, 1953–Sept. 17, 1957.

286. Voltage Regulator, N. Silver. *No. 2,809,341*, Feb. 13, 1956–Oct. 8, 1957.

287. Voltage Regulator, N. Silver. *No. 2,810,877*, Apr. 10, 1956–Oct. 22, 1957.

288. Magnetic Amplifier, E. L. Gardner. *No. 2,812,391*, July 6, 1954–Nov. 5, 1957.

289. Servo System with Magnetic Rate-Taking, H. Hecht, P. G. Degraff. *No. 2,812,487*, Dec. 30, 1953–Nov. 5, 1957.

290. Voltage Regulator, J. S. Comins, P. J. Gallagher. *No. 2,814,773*, Feb. 4, 1957–Nov. 26, 1957.

291. Regulated D.C. Power Supply, D. G. Scorgie. *No. 2,816,260*, July 31, 1956–Dec. 10, 1957.

292. Induction Motor Control Systems, E. C. Rhyne, Jr. *No. 2,817,051*, Dec. 22, 1956–Dec. 17, 1957.

293. Power Transformer Circuit, W. R. Oppen. *No. 2,817,770*, Apr. 17, 1956 –Dec. 24, 1957.

C. Magnetic Frequency Conversion

1957

294. Magnetic Frequency Divider, M. W. Smith. *No. 2,787,755*, Aug. 13, 1953–Apr. 2, 1957.

295. Magnetic Device, Shou-Hsien Chow. *No. 2,794,164*, Feb. 20, 1953–May 28, 1957.

D. Operational and Instrument Magnetic Amplifiers

1957

296. Magnetic Modulator System, G. Wennerberg. *No. 2,790,948*, June 19, 1951–Apr. 30, 1957.

E. Static Magnetic Switching Applications

June 1956 to December 1956

297. Magnetic Amplifier Over-speed Trip Devices, W. F. Horton, H. J. Tyler. *No. 2,760,149*, May 17, 1952–Aug. 21, 1956.

298. Magnetic Coincidence Detector, R. A. Ramey, Jr. *No. 2,766,420*, Mar. 16, 1953–Oct. 9, 1956.

299. Magnetic Delay Line, R. A. Ramey, Jr. *No. 2,770,737*, May 18, 1953–Nov. 13, 1956.

300. Trigger Circuit of the Magnetic or Dielectric Type, S. Duinker. *No. 2,773,198*, July 23, 1953–Dec. 4, 1956.

1957

301. Forcible Reversion of Magnetic Amplifiers, J. P. Eckert, Jr. *No. 2,792,507*, July 27, 1955–May 14, 1957.

302. Magnetic Amplifier Converter and Limiter, H. W. Kunnes, F. J. Antonazzi. *No. 2,796,535*, Sept. 30, 1954–June 18, 1957.

303. Bi-Polarity Relay Amplifier, H. W. Patton. *No. 2,802,982*, Dec. 15, 1955–Aug. 13, 1957.

304. Magnetic Amplifier Circuits, H. W Collins, R. W. Roberts, W. J. Dunnet. *No. 2,807,006*, Oct. 24, 1955–Sept. 17, 1957.

305. Bi-Directional Parallel Magnetic Amplifier, J. D. Lawrence, Jr. *No. 2,809,302*, Mar. 17, 1955–Oct. 8, 1957.

306. Saturable Core Reactor System, U. Lamm. *No. 2,818,545*, Sept. 20, 1952–Dec. 31, 1957.

F. Transistor Magnetic Device Combinations

July 1956 to December 1956

307. Transistor and Electromagnetic Control Apparatus, B. H. Hamilton. *No. 2,759,142*, Apr. 7, 1953–Aug. 14, 1956.

308. Magnetic Amplifier, R. L. Bright. *No. 2,773,132*, June 2, 1954–Dec. 4, 1956.

1957

309. Regulated Rectifying Apparatus, B. H. Hamilton. *No. 2,790,127*, Feb. 26, 1954–Apr. 23, 1957.

310. Voltage Regulator, J. S. Comins, P. J. Gallagher. *No. 2,801,383*, Sept. 24, 1956–July 30, 1957.

311. Polarity-Responsive Voltage Computing Means, R. L. VanAllen. *No. 2,808,990*, Oct. 31, 1956–Oct. 8, 1957.

312. Amplifiers, G. F. Pittman, Jr. *No. 2,809,343*, Dec. 24, 1953–Oct. 8, 1957.

IV: Foreign Patents

A. Theory, Analysis, and Design

1956

313. Magnetic Amplifier, B. D. Bedford. *German No. 941,494*, July 30, 1953–Apr. 12, 1956 (Priority U.S.A.: July 29, 1952).

D. Operational and Instrument Magnetic Amplifiers

1956

314. Direct Current Instrument Transformer, H. Gugel. *German No. 944,804*, June 10, 1943–June 21, 1956.

Reversible-Polarity D-C Power Amplifier Using Magnetic-Amplifier-Controlled Switched Transistors

N. L. SCHMITZ
MEMBER AIEE

T. BERNSTEIN
MEMBER AIEE

Synopsis: This paper describes, analyzes, and presents performance data on a reversible-polarity d-c power amplifier suitable for reversing drive of d-c motors in sizes up to 1 horsepower (limited only by presently available components) as well as for a wide range of additional applications including servomechanisms. Full-load and light-load efficiencies, output impedance, power gain, and other characteristics compare favorably with those of the rotating-amplifier-type motor-generator sets. The system under discussion, being completely static, can be expected to show significant advantages in reliability and ability to withstand a wide range of adverse environments. Limitations on size are practical rather than theoretical and may be expected to become less stringent as transistor technology improves and as further developments occur in areas such as switched silicon diodes.

THE PROBLEM of devising a high-efficiency magnetic-amplifier circuit with a reversible-polarity d-c output has been considered in some detail in the literature. Conventional reversible-polarity circuits[1,2] have a maximum theoretical efficiency of 50% as compared to 100% for the circuit to be considered. The increase in efficiency is accomplished by using switched transistors in conjunction with magnetic amplifiers. Previous applications of switched transistors[3] have had magnetic-amplifier gate windings in series with the transistors. Such circuits have the disadvantage of requiring the use of power magnetic amplifiers.

The advantages of operating transistors in their switching mode are well known.[4] These advantages include small power loss and consequent low temperature rise. Since the output of a self-saturating magnetic amplifier is a switched form of sine wave, it is well suited for control of transistors in their switching mode. In such a scheme the transistors control the load power while the magnetic-amplifier output power need only be sufficient to control the transistors. The use of switching transistors controlled by magnetic amplifiers rather than an all-transistor circuit can be justified in that the most important advantages of magnetic-amplifier control such as isolated, multiple inputs are retained.

Basic Circuit

Fig. 1 is a schematic representation of the basic circuit for providing reversible-polarity direct current for a 2-terminal load. The magnetic amplifiers are designed for d-c output and are biased to cutoff. Ideally, there is no base drive in the transistors under quiescent conditions and both transistors are nonconducting.

In actual operation, however, the exciting current required by the magnetic-amplifier cores at cutoff prevents this condition from being realized completely. Failure to render both transistors nonconducting under quiescent conditions results in an undesirably large circulating current and a balancing problem.[5] This effect can be minimized as outlined in Appendix I.

The control windings of the magnetic amplifiers are so interconnected that one polarity of control voltage will aid the bias on one amplifier and oppose the bias on the second. Reversed polarity of control voltage reverses this action. With a given polarity of control, the output of one magnetic amplifier provides base drive to the transistor and causes it to be switched on at a phase angle dependent on the magnitude of control signal. This switched output of the transistor then appears across the load.

Since the transistors are switched either fully off or fully on there is no gain-balancing problem associated with the use of the transistors. Based on the rms value of the output voltage the maximum theoretical efficiency is 100%.

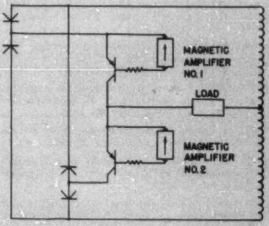


Fig. 1. Basic circuit for reversible-polarity d-c power amplifier

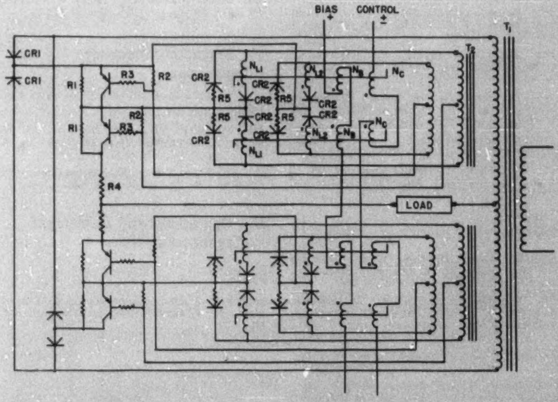


Fig. 2. Detailed circuit for reversible-polarity d-c power amplifier

---

Paper 58-1177, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 19, 1958; made available for printing July 31, 1958.

N. L. Schmitz and T. Bernstein are with the University of Wisconsin, Madison, Wis.

Vol.
Pt. 1
195

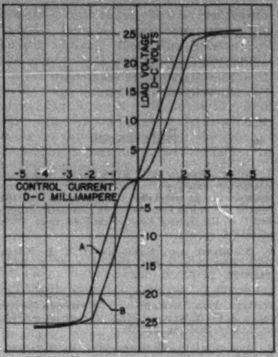


Fig. 3. Transfer characteristics for circuit of Fig. 2 with 10-ohm load

A—Excessive bias, 6.5 milliamperes
B—Normal bias, 6.0 milliamperes

Table I. Test Results

	10-Ohm Resistor Load				Motor Load	
Load voltage, d-c volts	25.5	20	12.5	0	30	15
Load current, d-c amperes	2.6	2.0	1.25	0	0.85	0.65
Load power, calculated from direct voltage and current	65	40	15.65	0	25.5	9.75
Load power, measured by wattmeter	87	69	38	0	27.5	10.5
Magnetic-amplifier power, measured by wattmeter	8.5	7	4.5	1.5	7.5	2.5
Power input to load circuit, measured by wattmeter	102.5	80	45	10	31	15
Efficiency, %, based on average current and voltage	58.6	46	31.6		66.4	55.8
Efficiency, %, based on measured power	78.5	79.4	76.6		71.5	60

Reversible-Polarity Circuit

Using the magnetic amplifier and transistor circuit of Fig. 1, the experimental circuit shown in Fig. 2 was developed. In order to provide a higher load voltage, two transistors are cascaded for each polarity of output. With transistors having suitably high voltage ratings, only one is required for each polarity of output. The 10,000-ohm resistors, R_1, across each transistor insure good voltage division. Additional transistors can be cascaded with the addition of load windings to the magnetic-amplifier core. The magnetic amplifiers are those discussed in Appendixes I and II.

The transfer characteristics (Fig. 3) of output voltage versus control current are dependent on the bias current. Too much bias current, shown by curve *A*, produces a dead zone about the zero control-current operating point. Too little bias current causes a large circulating current. With an ideal bias as shown by curve *B* there is some circulating current but it is not excessive. The 0.5-ohm resistors, R_4, of Fig. 2 tend to limit this current.

Experimental Results

Efficiency

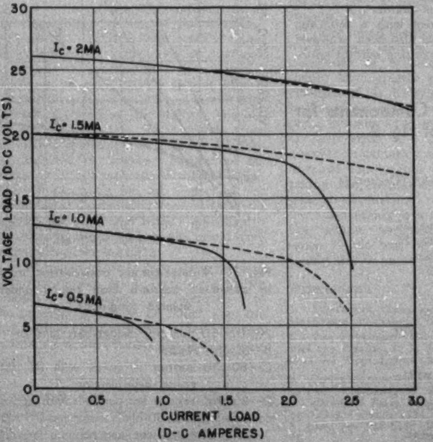
The efficiency of the circuit of Fig. 2 was measured using the components itemized in Appendix II.

Comparisons as to efficiency require a definition of terms. If a lossless circuit has a perfect full-wave rectified output, its output based on the product of average current and voltage is only 80.9% of the a-c power input. A wattmeter indication of load power includes losses caused by the a-c components of voltage and current in addition to the product of average voltage and current. In applications where the load is the armature of a motor with a separately excited field, it would seem that the useful output to be considered is that contributed only by the average components of the voltage and current. Table I lists the test results obtained using the circuit of Fig. 2.

Regulation

The problem of regulation of the output voltage under changing load becomes quite important when the load is a separately excited shunt motor. Under starting conditions the load current is large. Unless accompanied by sufficient base current, large load currents can cause the transistor to operate in the nonswitching mode. This mode of operation results in increased emitter-to-collector voltages which can cause excessive voltage regulation and excessive power dissipation in the transistor.

Fig. 4 shows the effect of base current



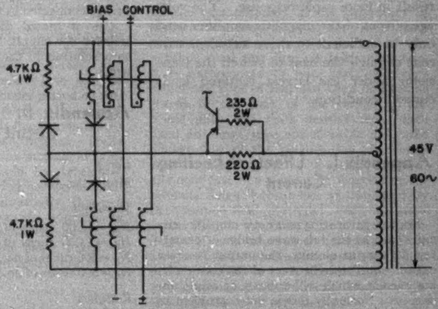


Fig. 4 (left). Voltage regulation for cicruit of Fig. 2

—— R_2 = 220 ohms, R_3 = 235 ohms
- - - - R_2 = 80 ohms, R_3 = 80 ohms

Fig. 5 (above). Detailed magnetic amplifier circuit for reducing effects of exciting current

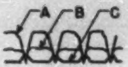


Fig. 6. Collect current with zero control signal

A—Before corrective measures
B—With shunt resistor only
C—With shunt resistor and balancing bridge

level upon voltage regulation. The sharp dropoff appearing in these characteristics results from the transistor being driven out of its switching mode as load current is increased. This effect can be minimized as shown by the dashed curves, by reducing the base circuit resistance and thereby increasing the base drive. Dropoff is still apparent at late firing angles (small control currents) because of the high peak currents required for a given average value of current.

Conclusions

The circuit of Fig. 2 lends itself well to the use of existing components. For higher current output, transistors can be paralleled without requiring additional magnetic amplifiers. For higher voltage applications transistors can be cascaded as shown at the expense of additional load windings on the magnetic amplifiers and on the small magnetic-amplifier supply transformer.

In the design of the magnetic amplifiers, the cores should be as small as possible. This reduces the exciting current and control power. The smaller cores also result in more rapid response. The voltage output of the magnetic amplifier need only be sufficient to drive sufficient current through the base to switch the transistor under the largest required load-current condition.

Appendix I. Effects of Exciting Current

For self-saturating magnetic-amplifier circuits such as the full-wave bridge or center-tap d-c output circuits, the output is never cut off completely because of the core exciting current which still appears at minimum output. Normally this is not a problem in a magnetic-amplifier circuit since the exciting current is usually quite small compared to the outputs desired. When this exciting current appears as base drive for a grounded-emitter transistor, however, it is multiplied by a factor having the order of 100 for large power transistors and appears in the collector load circuit. This causes a circulating current through the transistors even though the magnetic amplifiers are cut off.

Such a circulating current has two major disadvantages:

1. Unbalance in the gain of the transistors causes a voltage to appear on the load with no control signal.

2. The transistors are not operating in a switching mode since they are not completely cut off. Appreciable transistor heating is thus caused by the circulating current.

The exciting current problem can be minimized by using the smallest magnetic-amplifier cores possible and by using core materials such as Hy Mu-80 or Permalloy, both of which require very low exciting currents.

A further step towards alleviating the exciting current problem is shown in Fig. 5. The basic magnetic amplifier in Fig. 5 is a center-tap d-c magnetic amplifier providing base drive for the transistor through the 235-ohm resistor. With zero control signal, the exciting current of the magnetic amplifier appears in the 235-ohm base resistor. The emitter-to-base current can be reduced by shunting part of the exciting current through the 220-ohm resistor as shown. The magnetic amplifier must be designed to provide sufficient output current to drive the transistor into its "switched-on" mode with the parallel resistor shunting part of the base drive.

To reduce the collector current still further, the two diodes and 4,700-ohm resistors are added in Fig. 5 to form a bridge circuit with the 4,700-ohm resistors chosen so as approximately to equal the unsaturated impedance of the cores. This further addition has negligible effect on the normal operation of the circuit but does reduce the exciting current to a condition under which the exciting current is no longer a problem. Fig. 6 shows the three conditions for: (1) collector current caused by the exciting current (with no shunting resistor or balancing bridge); (2) shunting resistor alone; and (3) with the shunting resistor and balancing bridge connected in the circuit.

Use of biased diodes with a separate, isolated d-c supply has also been proposed as a solution to the exciting current problem.[5]

Appendix II. Components for Circuit of Fig. 2

Cores

Magnetics Inc. *51031-2A* (Orthonol)

Windings

N_{L1}=1,000 turns of no. *25*
N_{L2}=1,000 turns of no. *25*
N_B=500 turns of no. *25*
N_C=500 turns of no. *25*

Rectifiers

CR_1=transitron *TR 301*
CR_2=General Electric *1N91*

Resistors

R_1=10,000 ohms, 1/2 watt
R_2=220 ohms, 2 watts
R_3=235 ohms, 2 watts
R_4=0.5 ohm, 5 watts
R_5=4,700 ohms, 1/2 watt

Transistors

Delco *DT-100*

Transformers

T_1=115 volts/70 volts center tap
T_2=115 volts/45 volts center tap

References

1. HIGH EFFICIENCY PUSH-PULL MAGNETIC AMPLIFIER AND ITS USE AS A D-C SHUNT MOTOR DRIVE, C. L. Boyajian. "Proceedings of the Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T-98*, Sept. 1957, pp. 35–44.

2. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, chap. 26.

3. HIGH-EFFICIENCY PUSH-PULL MAGNETIC AMPLIFIERS WITH TRANSISTORS AS SWITCHED RECTIFIERS, A. G. Milnes. *AIEE Transactions*, vol. 77, pt. I, July 1958, pp. 327–31.

4. JUNCTION TRANSISTORS USED AS SWITCHES, R. L. Bright. *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

5. MAGNETIC AMPLIFIER CONTROL OF SWITCHING TRANSISTORS, H. William Collins. *Ibid.*, vol. 75, pt. I, Nov. 1956, pp. 585–89.

◆

Discussion

M. G. Rekoff, Jr. (Carl C. Crane, Inc., Madison, Wis.): The authors have made an interesting contribution to the rapidly expanding field of switched transistors with this paper. There is one point that is

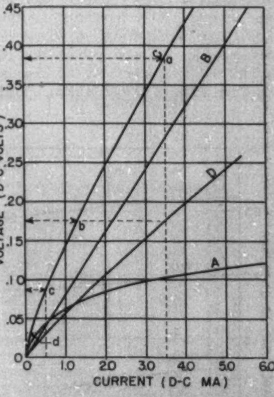


Fig. 7. Voltage-current characteristic used to determine transistor base current under shunted conditions

A—DT-100 transistor emitter-base junction
B—80-ohm resistor
C—80-ohm resistor in series with DT-100 emitter-base junction
D—80-ohm resistor in parallel with series combination of an 80-ohm resistor and DT-100 emitter-base junction

not clear; that is the effect of the emitter-to-base shunting resistance upon the emitter current. Since the resistance of the emitter-base junction is nonlinear the current will not divide directly proportional to the value of the base and shunting resistances. Fig. 6 seems to illustrate this effect since, although the shunting and base resistances are approximately equal, the collector current after shunting is less than half of the original current value.

Would the authors care to comment or elaborate on that facet of the circuit operation?

N. L. Schmitz and T. Bernstein: The effect of the emitter-to-base shunting resistor is difficult to analyze except by graphical means because of the nonlinear characteristics of the emitter-to-base junction. In Fig. 7, curve *A* represents the emitter-to-base current-versus-voltage characteristic for a typical *DT-100* germanium transistor while curve *B* represents an 80-ohm resistance characteristic. Using this emitter-to-base and resistance characteristic, the combined characteristic, *C*, of this junction characteristic in series with the 80-ohm resistor is constructed. The final characteristic, *D*, is constructed to show the current-versus-voltage characteristic of *C* in parallel with the 80-ohm resistor. This characteristic, *D*, now shows the plot of current versus voltage for the emitter-base junction in series with an 80-ohm resistor and the combination in parallel with an 80-ohm resistor. The exciting current output of the magnetic amplifier can be considered as a constant current source. The important component of this current is the current through the emitter-base junction since this current is multiplied by β, approximately 100, to obtain the collector current at minimum output.

If the size of the core and number of load turns were such as to provide 3.5 milliamperes of exciting current, then with no shunting resistor the current through the emitter base junction is 3.5 milliamperes as shown by point *a* on curve *C*. If there is an 80-ohm shunting resistor used then the exciting current through the junction is approximately 1.3 milliamperes as shown by point *b* of curve *C*. A lower shunting resistor might be used but then at maximum output more than half the control current would be diverted around the junction with a consequent loss in efficiency. For low exciting currents, for example 500 microamperes, the reduction in exciting current through the junction is more pronounced when shunting resistance is used because of the nonlinearity of the curves at low currents. This is shown at points *c* and *d* on curve *C*.

Another approach to the exciting current problem suggested by Prof. R. A. Greiner of the University of Wisconsin is to shunt only the emitter-base junction with a resistor rather than shunting the junction in series with a base resistor. By a graphical construction similar to the methods used in Fig. 7, shunting the emitter-base junction with a 10-ohm resistor while maintaining an 80-ohm resistor in the base circuit will reduce a 3.5-milliampere emitter-base current to approximately 0.25 milliampere. This result was also verified experimentally. The disadvantage of this method is that less current is driven through the emitter-to-base junction when the magnetic amplifier fires.

LATE DISCUSSIONS

Discussion of The Ferroresonant Circuit, George E. Kelly, Jr., pages 843–48.

C. F. Spitzer (Lockheed Aircraft Corporation, Palo Alto, Calif.): I concur fully with the author's statement that better agreement between analytically predicted results and experiment will be derived from a careful consideration of the harmonic components of current. This is, of course, primarily true for operation in the high-current condition, in which the inductor undergoes saturation in every cycle. Some simplified approaches to this problem have been discussed in the literature,[1,2] and a direct application to the simple resistance-inductance-capacitance version of the ferroresonant electric circuit is very readily made. An experimental confirmation of this need to include the harmonics is found in the difference between calculated and experimental values for essentially all circuit configurations tested by Mr. Kelly. The improved approximation which results from the neglect of circuit resistance can possibly be explained as follows: Neglect of resistance tends to raise the calculated value, thus compensating for the neglect of the harmonics of current which would also tend to raise the calculated value. Needless to say, considerations of harmonics would have been much closer to reality.

In its review of the history of the problem of ferroresonance the author refers to the Odessey-Weber paper (see reference 5 of the paper). It should be noted that the first method Mr. Kelly uses is identical with equation 2 in the Odessey and Weber treatment:

$$E=\sqrt{(IR)^2+\left(E_2-\frac{1}{\omega C}\right)^2}$$

although a comparison of experimental with analytically obtained results is not made there. However, the following statement follows this equation:

> A refinement of this method, due to Rouelle (see reference 3 of this discussion) requires the separation of the reactor voltage into active and reactive components on the basis of fundamental frequency. A predetermination of the operating characteristics shows good agreement with experimental results.

It thus appears from this description that Rouelle used to good advantage the second method described by Mr. Kelly, whose results appear to confirm those obtained by Rouelle.

I have found the fleeting reference to the ferroresonant state of a 3-phase 4-wire Y-connection of considerable interest, and should have greatly preferred to see the emphasis of Mr. Kelly's paper placed on this topic and on related problems of 2- and 3-phase systems with open lines, unbalanced or open loads, and open neutral. Quite possibly an extension of the method of symmetrical components could be developed, taking due account of the harmonic components of currents. Admittedly, this is a more demanding task than the approach to the simple resistance-inductance-capacitance circuit but, by the same token, it should also be much more rewarding.

REFERENCES

1. NONLINEAR VIBRATIONS (book), J. J. Stoker. Interscience Publishers, New York, N. Y., 1950.

2. THE FERRORESONANT TRIGGER PAIR: ANALYSIS AND DESIGN, C. F. Spitzer. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 407–16.

3. CONTRIBUTION À L'ETUDE EXPÉRIMENTALE DE LA FERRORÉSONANCE, E. Rouelle. *Revue Générale de l'Électricité*, Paris, France, pt. I, Nov. 24, 1934, pp. 715–41; Dec. 1, 1934, pp. 736–80; pt. II, Dec. 8, 1934, pp. 785–819; Dec. 15, 1934, pp. 841–58.

Discussion of Radio Attenuation at 11 Kmc and Some Implications Affecting Relay System Engineering, S. D. Hathaway, H. W. Evans, pages 930–38.

H. C. Franke (Bell Telephone Laboratories, Inc., Whippany, N. J.): This paper is one of the first attempts to present path performance suitable for engineering of microwave radio systems operating in the 11-kmc region.

The information which is presented should add to the wealth of data already available in other regions of the microwave spectrum and thereby aid the system engineer in further extrapolating on the suitability of higher frequency ranges for radio-relay communication.

While these test data are fairly complete, it is unfortunate to note that one very important type of test had been omitted as stated in the text of the paper. Additional path loss tests besides those with fixed antennas give considerable increased background of the propagation characteristics over the path. These tests are usually conducted by a series of observations of antenna height versus loss as measured by the received field strength. Such measurements include considerations of the relative height of each antenna as determined from an examination of the path geometry which includes height of each site, clearance over the intervening obstructions, and the desired clearance over the path. For an experiment of this type as many related factors as possible should be explored so that more complete data are available which should lead to better understanding of the effects of site selection, path clearance, effect of vegetation, etc. It is generally conceded that, in so far as possible, reflections on microwave paths should be avoided, since this secondary effect may increase the depth of fade and thereby result in poorer performance. Generally, from an examination of the path topography, tentative conclusions may be reached as to the possibility of reflections; particularly conducive to such conditions are flat barren areas or large bodies of water. While the test area appears to have considerable surface roughness when related to the wavelength of the frequency used, in this case equal to about one inch, the possibility of side reflections from man-made and natural smooth obstructions located in the path should not be ruled out.

While the attenuation due to rain is severe at these frequencies as pointed out in the data, it should be noted that during these periods the air is generally well mixed, and the possibility of selective fading due to reflections from atmospheric discontinuities is relatively small. Hence at any given time either fading loss due to multipath transmission or attenuation due to rain loss is present, but not both simultaneously. Therefore, the equipment margins need only consider the more severe condition.

◆

Discussion of Reactions of Bistable B-H Loops on Magnetic Amplifiers, R. E. Morgan, Communication and Electronics, November 1958, pages 586–90.

T. Bernstein and **K. Mei** (University of Wisconsin, Madison, Wis.): Mr. Morgan's discussion of the possible connection between the bistable **B-H** loops of the core and the instability of magnetic amplifiers at reduced voltage is an interesting study concerning an old problem. It seemed worth while to correlate this connection by investigating various materials to see if the instability occurs only when bistable **B-H** loops are evident.

Along this line of investigation, four samples each of 2-mil squared permalloy, permalloy, Deltamax, and L Silectron were wound and connected in a doubler circuit with a-c output. The cores used were Arnold Engineering *1T4180* with load and control windings each of 1,000 turns American wire gauge no. *25*. The excitation frequency was 60 cycles per second. Examining the flux-current loops of the cores alone after a d-c bias had been applied and removed, it was observed that all the Deltamax cores exhibited the bistable **B-H** loops while none of the other cores did. Operating the doubler circuit at 10% below the cores' saturation voltage, a criterion was established where the voltage output across the load resistor after the snap-up was compared to the zero control current voltage output. For control, a d-c power supply in series with a 10,000-ohm resistor was used while the load resistor was 100 ohms. Silicon diodes were used for the rectifiers.

Table I

Core Material	Voltage After Snap as Percentage of Zero Control
Squared permalloy	62.5
Permalloy	no snap
Deltamax	16.7
L Silectron	95

The results were as shown in Table I.

The interesting result is the lack of correlation between bistable loops and the instability. The results with the Silectron are particularly surprising. Has the author investigated any other core materials in the pulse-type circuit to see whether there is still instability even though there are no bistable **B-H** loops?

The author states, "From 12 cores selected at random, 8 of the cores have **B-H** loops (exhibiting bistable characteristics) while the remaining 4 cores have symmetrical minor **B-H** loops . . ." Were the four cores used in the circuit to see if they also caused instability? It would seem that cores which have more than two stable minor loops might exhibit more than one region of instability. Have such results ever been observed?